Vol. XVI Part 1

May 1963

THE BRITISH JOURNAL OF STATISTICAL PSYCHOLOGY

EDITED BY
R. J. AUDLEY

WITH THE ASSISTANCE OF
CYRIL BURT

AND THE FOLLOWING EDITORIAL BOARD

A. C. AITKEN
M. S. BARTLETT
W. G. EMMETT
M. G. KENDALL
D. N. LAWLEY
E. S. PEARSON
E. A. PEEL
L. S. PENROSE
J. FRASER ROBERTS
A. RODGER
W. STEPHENSON
A. STUART
P. E. VERNON

Printed and Published by
TAYLOR & FRANCIS LTD.
RED LION COURT, FLEET STREET, LONDON, E.C.4

Price 20*s. per part* (*U.S.A.* $3.50)   *Subscription* 30*s*. 6*d*. *per volume* (*U.S.A.* $4.50)

EDNL. & PSYL.

PUBLICATIONS OF THE BRITISH PSYCHOLOGICAL SOCIETY

The British Journal of Statistical Psychology is issued by the British Psychological Society. The subscription price per volume is 30*s*. net for Members of the Society and 30*s*. 6*d*. (post free) for non-members. The subscription price in the U.S.A. is $4.50 (post free). Members of the Society should send their subscriptions to THE SECRETARY, British Psychological Society, Tavistock House South, Tavistock Square, London, W.C.1 : non-members to Messrs. TAYLOR & FRANCIS, LTD., 18 Red Lion Court, Fleet Street, London, E.C.4. Until further notice the Journal will be issued in two parts each year, about May and November.

Papers for publication and books for review should be sent to the Assistant Editor, Sir Cyril Burt, 9 Elsworthy Road, Hampstead, London, N.W.3. Authors should retain copies of their manuscripts as neither the Society nor its Editors can undertake any responsibility for loss or delay in the case of unsolicited contributions. Authors will receive 50 copies of their articles free; extra copies may be purchased provided the order is given when the proofs are returned. Contributors should indicate whether they wish their reprints to be bound in covers. The printers will state current prices when sending out proofs.

The Society also issues quarterly *The British Journal of Psychology* and *The British Journal of Medical Psychology*, at a subscription rate of 60*s*. per annum for either journal, and *The British Journal of Social and Clinical Psychology* three times a year at a subscription rate of 50*s*. per annum. Subscriptions for these journals should be sent to the Cambridge University Press, Bentley House, Euston Road, London, N.W.1. The Society publishes, jointly with the Association of Teachers in Colleges and Departments of Education, *The British Journal of Educational Psychology* three times a year: for this the subscription, 30*s*. per annum, should be sent to Methuen & Co. Ltd., 36 Essex Street, London, W.C.2. Members of the Society receive the foregoing journals on special terms; enquiries should be addressed to the Secretary, British Psychological Society, Tavistock House South, Tavistock Square, London, W.C.1.

Papers for publication in *The British Journal of Psychology* should be sent to Dr. Boris Semeonoff, Department of Psychology, The University, Edinburgh 8, those for publication in *The British Journal of Medical Psychology* to Dr. Thomas Freeman, Lansdowne Clinic, 4 Royal Crescent, Kelvingrove, Glasgow, C.3; those for publication in *The British Journal of Social and Clinical Psychology* to Mr. Michael Argyle (Social Psychology Editor), Institute of Experimental Psychology, 1 South Parks Road, Oxford, or to Dr. Jack Tizard (Clinical Psychology Editor), Maudsley Hospital, Denmark Hill, London, S.E.5; and those for publication in *The British Journal of Educational Psychology* to Mr. L. B. Birch, University of Sheffield, Institute of Education, Sheffield 10.

PREPARATION OF PAPERS

Contributors are asked to send their papers in a form which is ready for submission to the printer: that is to say, the arrangement of the material should follow that observed in previous publications of this *Journal*. Each article should be headed either with a series of numbered headings in italic, corresponding with the cross-headings of the several sections, or with a brief abstract: (the latter procedure is preferred). Each section should have a short cross-heading, in capitals; and the whole should conclude with a summary embodying the main conclusions reached. Special attention should be paid to such details as correct spelling, grammar, capitalization, numbering, etc. (particularly in the case of tables and references). In preparing manuscripts and in correcting proofs authors should conform, so far as possible, with the 'Recommendations to Authors' set out in *The Printing of Mathematics* by T. W. Chaundy, P. R. Barrett, and Charles Batey (Oxford University Press, 1954, ch. II, pp. 21–73), or to those given in the paper on 'Setting Mathematics' (Arthur Phillips, *Monotype Recorder*, XL, iv, 1956).

Owing to the present conditions for printing, the cost of setting up tables and algebraic formulae which require hand composition is very much greater than the cost of machine composition. Contributors are therefore advised to arrange their material so that it can be printed as economically as possible. Often numerical results and algebraic formulae can, with a little simplification, be readily adapted for purposes of machine composition. Manuscripts involving many tables and numerous equations can as a rule only be accepted if the authors are willing to pay for the additional cost. Contributors should note that the *Journal* is intended for the general psychological reader, not solely for those who specialize in statistics. Unusual technicalities should be either avoided or carefully explained, e.g. by reference to previous discussions of the same or kindred problems in this *Journal*.

In future authors will receive one complete slip-proof of their papers, but no page proof. They will be expected to pay the cost of all corrections for which they themselves are responsible.

Vol. XVI
Part 1

The British Journal of Statistical Psychology

May
1963

AN EXPERIMENTAL STUDY OF VERBAL FLUENCY FACTORS[1]

32

By PAUL R. CHRISTENSEN and J. P. GUILFORD
System Development Corporation University of Southern California

The objective of the investigation here described was to carry out a factor-analytic study of the somewhat ambiguous area of verbal fluency, using a new approach which emphasized certain experimental features. Hypotheses regarding difference among the fluency factors were tested by systematic variations in sets of two or more parallel tests.

A six-hour battery of experimental and marker tests was administered to 221 Naval Air Cadets. Thirteen factors were extracted from the intercorrelations of 41 variables. First-approximation positions were obtained by placing each of nine reference axes near a single marker test in an orthogonal system. From this two rotational solutions were obtained graphically, one using the criteria of simple structure and positive manifold only with no knowledge of the tests, the other with knowledge and with the further criterion of psychological sense. The two were in general agreement.

The following factors were identified: (i) four reference factors, *Verbal Comprehension*, *General Reasoning*, *Writing Speed*, *Originality*, (ii) a factor called *Eduction of Conceptual Correlates;* (iii) three factors specific to parallel tests or overlapping scores; (iv) one residual factor, and (v) four fluency factors, namely *Expressional Fluency* (the ability to produce connected discourse, irrespective of the quality of production), *Ideational Fluency* (the ability to produce ideas in situations where there is little restriction and a high level of quality is not required), *Word Fluency* (the ability to produce words meeting structural requirements with moderate restrictions on responses), and *Associational Fluency* (the ability to produce isolated words to fit a given relation and unit of meaning).

Bureau of Educl. & Psyl. Research
(J. E. R. I.)

1. INTRODUCTION

Numerous investigators have studied verbal fluency by factor analysis, some restricting their attention to these factors only, others including them in broader studies which involve several classes of intellectual factor. From these studies a number of concepts have arisen which are believed to stand for basic dimensions of thinking or of personality. In comparing dimensions from one study with those in another, one can find some agreement, but also much disagreement as to the names and definitions of the concepts or the dimensions. Since the concepts remain more or less ambiguous despite the results, it was felt that a more searching approach should be adopted.

---

[1] Under Contract N6onr-23810 with the Office of Naval Research. Reproduction in whole or in part is permitted for any purpose of the United States Government. The opinions expressed are our own and are not necessarily shared by the Office of Naval Research.

Acc. J 851
Da

II. Fluency Factors from Previous Studies

In reviewing previous work it proved possible to recognize four, more or less persistent, verbal-fluency factors. Comparisons were made of studies bringing out two or more such factors and involving a rotation of axes. These criteria were employed to provide a basis for differentiation within the studies and identification across the studies. Ten of the published studies were thus qualified [1, 3, 6, 8, 9, 12, 13, 14, 15, 16],[1] and an additional unpublished study also contributed some light on the subject.[2]

Certain difficulties interfered with the cross-identifications. To represent each factor, similar tests, but not always the same tests, were used in different analyses. By 'similar' we mean that common psychological processes seemed to be involved. Extending the factor involvement to a wider variety of tests yields some indication as to the general nature of the ability represented; and an examination of tests consistently *not* loaded on a factor helped to sharpen the conception. The fact that in some analyses certain tests are loaded on two or more factors, suggests that in such cases there may be some confounding of factors.

Word Fluency. Factors recognized as *word fluency* in the studies cited are loaded with tests in which word structure is important rather than semantic or ideational content. The tasks consist of producing words, usually disconnected, in limited time, where the condition to be satisfied is some structural attribute, such as spelling. The word produced is thus merely a symbolic unit, not a meaningful concept. Another feature is restriction of the class within which the words are to be produced. Of the four fluency factors, this is least ambiguous in reports of previous studies. A question we attempted to answer is whether some degree of restriction is optimal in testing for *word fluency*.

Associational Fluency. At the time this study was planned, the leading hypotheses about the nature of *associational fluency* were as follows. It involves verbal meaning rather than word structure, i.e., the content is semantic rather than symbolic. It represents the ability to think of single words meeting specific requirements as to meaning. It had not been clear what kinds of meaningful relations (e.g., synonyms, classes or associational linkages) are best for measuring this factor, or whether it is essentially a multiple-association factor, and to what degree the task needs to be restricted.

Ideational Fluency. In tests associated with *ideational fluency* the score is the number of ideas produced in response to a theme, a question, or a picture. The idea may be expressible in one or more words. The responses need not be of high quality. In some of the tests loading previously on this factor, the clever or remotely connected responses are excluded in scoring. Furthermore, the number of high-quality responses has definitely lower loadings on this factor

---

[1] Considerably more detail concerning comparisons of these studies will be found in (7).

[2] An analysis performed by the Aptitudes Project at the University of Southern California on data kindly supplied by the Institute for Personality Assessment and Research, University of California at Berkeley.

[15, 8]. In testing it, therefore, there is little or no restriction in terms of the evaluation of responses.

One of the anomalies in earlier studies is the correlation of tasks involving naming members of a class with both *associational fluency* and *ideational fluency*: e.g., tests such as Things Round and Animals [13, 12]; in these the tasks are consistent with the characteristics of either factor.

Expressional Fluency. Tests of expressional fluency indicate that the ability most involved pertains to organizing words into phrases or sentences. A number of questions arise, however, in reviewing previous studies. Is high quality an important attribute? Since the Similes test has been a prominent measure of *expressional fluency* [3, 13, 14], is the factor most appropriately involved in turning out figures of speech? If so, is the important function that of producing attributes, objects, or expressions? Is there a relation between number of connected words required and the correlation with the factor? These and other questions are embodied in the research design described in the next section.

III. THE RESEARCH DESIGN

A. *General Plan and Preliminary Analyses*. The general plan included (1) hypothesizing the separate existence of four factors of verbal fluency, as just described; (2) hypothesizing in what respects these four factors differ from one another; (3) developing for each of these differences two or more tests to differ systematically with regard to a selected variation, e.g., degree of restriction in response category; (4) selection or development of marker tests; (5) pre-testing; (6) eliminating unpromising hypotheses and revising tests; (7) administering experimental and marker tests; (8) factor-analysing the scores; and (9) reconsidering the hypotheses in the light of the results.

The preliminary testing served several functions. One was in deriving item statistics, which made possible the development of comparable tests that differ along a single dimension. Another was in developing parallel parts of the same test in order to obtain good estimates of parallel-forms reliability. Pretesting also led to better test instructions, and provided opportunity to develop scoring procedures.

Eleven one-hour batteries were pretested on groups of college students.[1] All items (each usually calling for multiple responses) were separately timed. From statistical information, including item intercorrelations and test intercorrelations, unpromising hypotheses and variations were eliminated. This was in recognition of the need to keep the total battery within the available testing time. The question 'Are these test scores likely to differ considerably in factor content?' was constantly kept in mind.

---

[1] We are indebted to the following persons at the University of Southern California for making classes available for preliminary testing: Drs. Constance D. Lovell and Everett J. Wyers, Psychology; Drs. Edward C. McDonagh and Melvin J. Vincent, Sociology; and Mr. Robert A. Jones, School of Education.

B. *Test Variations and Related Hypotheses.* Hypotheses in this study are concerned mainly with the ways in which the factors of fluency differ from one another and from other factors. The hypotheses will be considered in relation to each of the systematic variations in tests or the conditions of testing.

1. *Variation in Temporal Period of Production.* When an individual produces multiple responses to fit a specified class, his rate of production is ordinarily high in the beginning but progressively slower as time elapses [2]. Apparently the early production of words is indicative of a personal quality of rate of evocation, while later production is indicative of the personal supply of appropriate words. It is suggested that there may be a difference in factor content of test scores, according as a score indicates rate of output or available supply. To study this relation several scores were derived from Controlled Associations, in which the examinee had to write as many synonyms as possible for each word. Variation in temporal period was made with four test scores as follows (test numbers correspond with those in Tables 2–3): 7. Controlled Associations IIIa—Synonyms written during first ½ minute; 8. Controlled Associations IIIb, c—Synonyms written during the next 1½ minutes. 9. Controlled Associations IIId—Synonyms written during the third minute. 10. Controlled Associations IIIe—Synonyms written during the fourth and fifth minutes.
Hypothesis: The earlier the period of production, the greater the variance in *ideational fluency*; the later the period of production, the greater the variance in *originality*; for intermediate times a maximum variance in *associational fluency*. The hypothesis assumes that in this kind of task early production results in responses relatively easy to call up, and that the late period requires the examinee to search for uncommon and remote associations; hence the predictions as to the relative importance of *ideational fluency* and *originality*. In between the more easily evoked responses and the more difficult there should be an optimal measure of *associational fluency*.

2. *Variation in Degree of Restriction of Class.* Since tests of the thing-listing type have been found to correlate sometimes with *ideational fluency*, sometimes with *associational fluency*, and more often perhaps with both, there seemed to be a real advantage in varying several tests in regard to restriction. Three tests were revised to represent three levels of restriction in terms of breadth of the class. Empirical evidence for effects of this variation should lie in the magnitude of the mean number of responses produced per minute.

16. Thing Listing III—high restriction.
17. Thing Listing II—medium restriction.
18. Thing Listing I—low restriction.

Hypothesis: The greater the class restriction, the greater the variance in *associational fluency*; the less the restriction, the greater the variance in *ideational* *fluency*. There are at least two sources of restriction among this trio of tests—number and kind of attributes specified. In tests 16 and 17 three attributes are specified for the things to be listed, and in test 18 only two.

3. *Variation in Number of Structural Specifications in Isolated Words.* The best measures of *word fluency* have been tests in which words are produced to meet certain structural requirements. The tasks are thus highly specific. This specificity can be stated in terms of the number of mechanical or structural conditions the responses must satisfy, as in the following tests:

3. Word Listing III—three letters given.
4. Word Listing II—two letters given.
5. Word Listing I—one letter given.
6. Word Listing O—no letters given.

Hypothesis: The greater the number of specified letters a response must contain, the higher should be the loading on *word fluency*; the smaller the number the higher the loading on *ideational fluency*. It was not anticipated that the loadings on *ideational fluency* would necessarily be high, only that they would vary in the predicted direction, since hardly any of the other requirements of an *ideational fluency* task are present in these four tests.

4. *Variation in Number of Connected Words.* If *expressional fluency* involves the ability to produce connected discourse, tests differing in this respect should be differentially related to the factor. And, since both *word fluency* and *associational fluency* correlate predominantly with tasks in which there is a production of *isolated words*, the closer a test approaches that condition, the more it should correlate with one or both of those factors. Accordingly, two groupings of tests were made: in the first three mentioned below they require the production of one word, two connected words, and four connected words, respectively. This progression is assumed to be along a dimension of increasing involvement with continuous discourse.

6. Word Listing O.
25. Two-Word Combinations.
27. Four-Word Combinations.

Hypothesis: The more words used together, the higher should be the loading on *expressional fluency*; the fewer words used, the higher the loadings on *word fluency* and *associational fluency*. The same hypothesis applies to the next set of tests. In these the task differs from the set above in that, instead of listing simply words or word combinations, there is the additional requirement that the first letter of every word used must be the letter specified.

5. Word Listing I.
26. Two-Word Combinations—First Letters Given.
28. Four-Word Combinations—First Letters Given.

5. *Variation in Number of Structural Specifications in Connected Discourse.* In several studies reported by French [5] there is a suggestion that theme-writing scores are a function of two main abilities, *expressional fluency* and *ideational fluency*. It would appear that for scores related to the total number of words used, the task set by the topic is so free as to depend heavily on ideational ability, at the same time requiring only enough expressive ability to make the ideas reasonably coherent. Decreasing the freedom, and making the requirements in the task structural rather than meaningful or ideational, might conceivably effect a reduction in *ideational fluency* and an increase in *expressional fluency*. The following tests were designed in parallel form to help discover the effect of varying structural restriction. The second of each pair of tests below requires that the given letters be used in each word making up the combination; the first has no such requirement.

25. Two-Word Combinations.
26. Two-Word Combinations: First Letters Given.
27. Four-Word Combinations.
28. Four-Word Combinations: First Letters Given.

Hypothesis: With no structural specifications there should be relatively more *ideational fluency*, and with the single element specified for each word there should be more *expressional fluency*.

6. *Variations in the Process of Constructing Similes.* Not all the features that may differentiate factors can be considered in quantitative terms. Qualitative differences were studied in connection with the simile-construction task. In the analyses by Carroll [3], Taylor [13], and Rogers [12], simile-completion tests appeared to be at least as good as any other for measuring *expressional fluency*. It seemed worth while to make certain variations in the simile-construction task with a view to increasing the *expressional fluency* variance and clarifying the differences between this factor and other fluency factors. The first variation to be considered is that of producing names of objects vs. attributes, both while completing similes. The psychological process would seem to be somewhat different in the two tasks. Test 23 requires calling up of a number of objects that have a given attribute in common with a given object; test 22 requires calling up of a number of attributes that two given objects have in common: e.g., 23. Simile Completion: ‘ The kitten's paws were as smooth as . . .’ 22. Simile insertion: ‘ The kitten's paws were as . . . as velvet.’ Hypothesis: Producing objects that have a given attribute in common with a given object should involve more *expressional fluency* variance; producing attributes that two given objects have in common should involve more *associational fluency* and *ideational fluency*. This hypothesis is based on the fact that the Simile Completion test is more like the test called Similes used in previous studies which was high in *expressional fluency*. The test in Taylor's study called Adjectives had its highest loading on *associational fluency* and *ideational fluency*.

A second comparison can be made between two simile tasks that exhibit another difference, namely, between producing a number of attributes that two given objects have in common (as in Simile Insertion) and composing a number of more or less complete expressions of the attributes that two given objects have in common (as in Simile Interpretation):

22. Simile Insertion: ' The kitten's paws were as . . . as velvet.'
24. Simile Interpretation: ' The kitten's paws were like velvet . . .'

One of the completions for the example from test 24 might be, ' because they had an exquisite smoothness '; a plausible attribute would suffice.
Hypothesis: Producing attributes that two objects have in common should involve more *associational fluency* and *ideational fluency*; producing multi-verbal expressions of the attributes should involve more *expressional fluency*. Logically the responses in the two tests answer the same question, namely, ' how are the two objects similar? ' We might therefore expect that the ideational processes would be involved to about the same degree. Where more single-word responses are possible, as in test 22, there should be more *ideational fluency*. Perhaps a more important difference is that test 24 requires the production of coherent discourse, which, according to one conception of the factor, should favour *expressional fluency*.

7. *Variation in Quality of Production.* Is high quality of response an important ingredient in *expressional fluency*? Variation between high and low quality is represented by two scores, each from a different test. Two from Plot Titles are the number of titles for a given story that are rated clever, and the number of titles rated as non-clever. For Word Synthesis there are also two scores: ten words are given which must be used in constructing a story; the quantity score is the total number of words used. The quality score is derived from ratings for the coherence of the story.

20. Plot Titles (low quality).
36. Plot Titles (cleverness).
29. Word Synthesis (words).
30. Word Synthesis (rating).

Hypothesis: Quantity scores or low-quality scores should correlate more with *ideational fluency*; high quality scores with *expressional fluency* (as well as with *originality*).

8. *Variation in the number of Connected Sentences.* How is *expressional fluency* loading affected by requiring the production of unrelated sentences vs. coherent *groups* of sentences? This was studied by comparing the Word Synthesis with Word Arrangement. The former requires the construction of a sequence of sentences coherently related; the latter does not. In each the

examinee is given a number of words to be used, ten in the former, four in the latter.

30. Word Synthesis: related sentences.
31. Word Arrangement: unrelated sentences.

Hypothesis: Producing unrelated sentences using specified words will result in more *expressional fluency*; producing coherent groups of sentences will result in more *ideational fluency*. This prediction arises from the belief that producing a sequence of sentences provides less restriction, a condition favourable for *ideational fluency*.

9. *Rearrangement of Words vs. Substitution of Words.* One way considered plausible for maximising the proportion of *expressional fluency* variance consists in requiring the examinee to express the same idea in different ways. Such a task can be presented in two ways: one calls for rearranging the elements of a sentence, the other for substituting elements. Since word substitution fits the *associational fluency* concept better, there is the possibility of a higher relation of *expressional fluency* to the rearranging activity. An arrangement score and a word-substitution score were derived from Alternate Headlines, which requires the production of different headlines, all containing essentially the same information as the headline given.

32. Alternate Headlines (arrangement).
33. Alternate Headlines (word substitution).

Hypothesis: Rearrangement of words should correlate more with *expressional fluency*; substitution of words more with *associational fluency*.

10. *Synonymous vs. Non-synonymous Association.* The factor of *verbal comprehension* often has some variance in tests of *associational fluency*. Tests of *verbal comprehension* commonly ask the examinee to select or suggest a synonym for the given word. *Associational fluency* is related to both synonymous and non-synonymous tests. A test was made of the expectation that this factor would be more related to non-synonymous tests. The variation was represented in the factored battery by tests both of which presented a high degree of restriction, calling for one unique answer. Both call for producing a word associated with two given stimulus words; in one case the response is a synonym to both, in the other it is not.

13. Associations III—synonymous.
14. Association IV—non-synonymous.

Hypothesis: Producing synonyms should favour *verbal comprehension* variance; producing other kinds of associations should favour *associational fluency* variance. A general statement should be made about the variations described in this section and the way the hypotheses are tested. The hypotheses are limited by the context of the tests representing them. This is stressed by stating the variations and hypotheses in operational terms. Another important

TABLE I. INFORMATION REGARDING TESTS.

Title	*Description*	*Estimated Reliability*[a]
1. Rhymes	Write words rhyming with a given word.	·74
2. Suffixes	Write words containing specific suffix.	·65[b]
3. Word Listing III	Write words containing three specified letters.	·78
4. Word Listing II	Write words containing two specified letters.	·70
5. Word Listing I	Write words containing one specified letter.	·74
6. Word Listing O	Write separate words, no restrictions.	·88
7. Controlled Associations—IIIa	Synonyms written during the first half minute	·87
8. Controlled Associations—IIIb, c	Synonyms written during the second half minute and the second minute.	·49
9. Controlled Associations—IIId	Synonyms written during the third minute.	·49
10. Controlled Associations—IIIe	Synonyms written during the fourth and fifth minutes.	·67
11. Vocabulary Completion	Produce a word that fits a given definition and begins with a given letter.	·81
12. Inventive Opposites	Write two antonyms for a given word, first letters given.	·71
13. Associations III	Produce a word that is synonymous with two given words.	·69
14. Associations IV	Produce a word that is not synonymous with but can be associated with two given words.	·62
15. Inventive Verbal Relations	Produce a word for each space that bears a specified relation to a given word.	·79
16. Thing Listing III	Write thing names fitting highly restricted classes.	·36
17. Thing Listing II	Write thing names fitting somewhat restricted classes.	·33
18. Thing Listing I	Write thing names fitting broad classes.	·40[b]
19. Brick Uses (fluency)	List different uses for a brick.	·59[b]
20. Plot Titles (low quality)	Write titles for story plots.	·85
21 Descriptive Completion	Write completions to simple sentences.	·73
22. Simile Insertions	Produce attributes that two given objects have in common.	·71
23. Simile Completion	Produce objects that have a given attribute in common.	·75
24. Simile Interpretations	Compose more or less complete expressions of the attributes two objects have in common.	·75
25. Two-Word Combinations	Write two-word phrases.	·84
26. Two-Word Combinations—First Letter	Write two-word phrases, first letter of each word given.	·73
27. Four-Word Combinations	Write four-word sentences.	·81
28. Four-Word Combinations—First Letter	Write four-word sentences, first letter of each word given.	·61
29. Word Synthesis (words)	Write a story containing 10 specified words.	·72
30. Word Synthesis (rating)	Score is rating on coherence of story.	·60
31. Word Arrangement	Write sentences containing four specified words.	·70
32. Alternate Headlines (arrangement)	Write alternate versions of given headlines. Score is number of changes in arrangement.	·49
33. Alternate Headlines (word)	Score is number of words substitutions.	·49
34 Verbal Relations Naming	Describe the relation that one word bears to another.	·75
35. Cartoons	Supply clever " punch lines " for cartoons.	·67
36. Plot Titles (cleverness)	Score is number of clever titles written.	·74
37. Brick Uses (flexibility)	Score is the number of times the class of uses is changed.	49[b]
38. Verbal Comprehension	Select the word that is similar in meaning to a given word.	·79
39. Marking Speed	Make X's in rows of boxes as rapidly as possible.	·85
40. Ship Destination	Find the distance from ship to port considering the influence of several variables.	·87[c]
41. Word-Group Naming	Give a class name to a group of five words.	·52

[a] All alternate-form estimates except where noted. [b] Communality as lower-bound estimate. [c] Kuder-Richardson, with Tucker modification.

limitation is that the range of a variation is limited by the tests representing it. The findings may be interpreted only within these limitations.

C. *Tests Used in the Study.*[1] The tests are set out in Table I. Parallel tests representing different experimental variations will be found in consecutive groups. Fourteen of the 41 tests have not been mentioned because they do not fit any of the variations discussed. They were included as marker tests for fluency and other factors. The factors and their markers are as follows:

Word fluency: 1, Rhymes; 2, Suffixes. *Ideational fluency*: 19, Brick Uses (fluency); 21, Descriptive Completion. *Associational fluency*: 11, Vocabulary Completion; 12, Inventive Opposites. *Verbal comprehension*: 38, Verbal Comprehension. *General reasoning*: 40, Ship Destination. *Spontaneous flexibility*: 37, Brick Uses (flexibility). *Originality*: 35, Cartoons. *Eduction of conceptual correlates*: 15, Inventive Verbal Relations; 34, Verbal Relations Naming. *Writing speed*: 39, Marking speed.

In a previous study (9) it was doubted whether the factor defined by 15 and 34 tests was *expressional fluency* or *eduction of conceptual correlates*. The liberal representation of the former offered an opportunity to resolve this issue.

IV. Test Administration and Treatment of the Data

A. *Test Administration and the Experimental Sample.* The tests were organized in six booklets, each requiring approximately 60 minutes. Similar tests were separated in different booklets. Nearly all the tests were divided into separately timed parts, care being taken regarding time limits. Examinees were told how much time they had for each part. The battery was administered to 230 Naval Air Cadets at Pensacola Naval Air Station[2] by Personnel at the U.S. Navy School of Aviation Medicine Psychological Laboratory. Motivation was reported as high, considering that the weather was hot and humid. Because complete data were not available for a few examinees, the number was reduced to 221.

B. *Treatment of Test Scores.*[3] *Scoring.* Except for some marker tests with answer sheets, the tests required the completion of responses. In scoring each response had to be judged according to its relevance. The general procedure was to prepare a protocol for scoring each test, on the basis of experience acquired in pretesting. Each was scored independently by two persons.

---

[1] We acknowledge permission to use or adapt tests by L. L. Thurstone (tests 2 and 12) and Frank Barron (test 29). Acknowledgement is also made to the *Saturday Evening Post* and *Colliers* for permission to use some of their cartoons in test 35.

[2] We are indebted to Dr. Wilse B. Webb for making this testing possible and to the staff of the Aviation Psychology Laboratory, U.S. Naval Air Station, Pensacola, Florida.

[3] Grateful acknowledgement is made to Dr. Norman W. Kettner who assisted in supervising the scoring of tests and the derivation of the transformation matrix for the analytical rotation.

Table II. Unrotated Factor Matrix†

	A	B	C	D	E	F	G	H	I	J	K	L	M	h^2
1.	57	32	16	28	09	−06	−19	−07	−09	−07	14	02	−06	62
2.	53	21	24	21	09	−24	−24	−10	11	04	−11	20	16	65
3.	54	46	21	24	06	−03	−15	−02	06	−02	−18	−13	07	69
4.	54	37	26	30	20	08	−09	07	−17	−04	−10	06	−07	70
5.	64	20	21	31	07	−05	−17	15	−12	05	09	−02	04	68
6.	53	−36	−03	26	−11	−22	11	13	06	09	02	08	12	60
7.	50	14	04	−03	−36	−10	−03	07	−12	22	14	18	−07	54
8.	49	16	37	−25	−30	−09	23	10	09	−03	−12	07	05	66
9.	47	07	24	−23	−28	−05	08	13	13	−14	09	−11	−07	50
10.	57	10	33	−31	−27	−17	15	06	09	−16	−07	10	−08	72
11.	62	28	−25	05	09	12	12	−18	11	09	14	05	−03	64
12.	57	29	−17	07	−11	09	12	−20	05	−06	21	−13	03	58
13.	51	42	−16	−05	15	−08	−17	−10	12	−13	09	20	10	62
14.	41	38	−11	−19	07	−21	−06	−07	04	−21	05	15	−11	50
15.	58	37	−28	−18	05	17	05	−12	−11	07	−08	−03	04	66
16.	44	−06	−10	−09	04	−08	02	16	−04	10	11	04	−10	28
17.	60	−19	05	−11	02	−16	03	−16	04	23	−13	−07	−13	56
18.	53	−09	−05	07	12	−04	06	−09	11	21	−12	−08	−06	40
19.	41	−38	22	−25	16	13	−07	10	27	15	−03	−02	−09	59
20.	09	−57	10	−16	−02	−15	−05	−08	04	13	−02	10	15	45
21.	56	−24	−04	09	19	−11	05	09	05	10	05	06	−10	47
22.	62	−09	07	−16	−15	16	−02	08	−06	16	17	−15	04	56
23.	51	−15	−06	−15	−14	−16	−15	05	−16	−10	−11	−06	10	44
24.	53	−37	09	−11	−09	24	−07	−24	08	−08	−03	−03	09	59
25.	57	−35	11	27	02	−07	31	08	−08	−11	04	−03	08	67
26.	60	−09	21	12	09	−04	14	−12	20	−02	07	07	08	53
27.	50	−35	−04	15	21	04	03	21	13	−24	08	−05	−05	57
28.	52	−19	11	21	06	18	14	−16	−13	−17	−10	16	−02	53
29.	36	−24	−21	23	−30	26	−30	18	09	11	−05	28	06	67
30.	37	11	−27	18	−33	33	−19	19	10	03	−14	−02	−08	58
31.	53	−38	12	−07	07	29	03	06	07	−11	−12	14	10	60
32.	22	−14	−19	−06	21	−22	−14	10	09	−11	05	−17	17	31
33.	40	−08	−03	21	−24	−12	20	−28	08	10	13	07	05	44
34.	62	27	−31	−12	16	14	06	08	−19	10	06	03	09	68
35.	33	−12	−28	−22	−17	−07	−25	−11	−12	−18	07	02	13	43
36.	65	−16	−04	−10	−09	−04	−08	−15	−18	−12	−07	−06	−18	59
37.	41	−21	20	−27	23	16	−03	05	03	−10	15	−16	15	49
38.	48	45	−15	−02	−10	07	17	−07	−02	08	−10	−11	−08	54
39.	13	−23	−19	10	−17	−13	06	10	−07	−09	−13	−08	−06	22
40.	32	23	−15	−14	26	04	18	06	−14	09	−18	−07	18	40
41.	40	17	−19	−12	21	08	12	−05	10	09	06	−03	−05	33

† Decimal points have been omitted.

TABLE III. ROTATED FACTOR MATRIX†

	A	B	C	D	E	F	G	H	I	J	K	L	M	h-
	VC	GR	MS	O	EF	IF	CC	WF	AF	Sb	Sc	Sw	R	
1.	30	−06	−03	07	13	−01	25	58	13	09	08	01	24	60
2.	24	05	01	09	06	18	10	71	19	−03	−02	07	−12	68
3.	52	07	01	−10	04	00	14	57	17	19	−02	−03	02	70
4.	39	16	−06	−09	16	−04	07	61	06	09	13	−02	27	70
5.	21	10	12	00	14	04	14	62	25	17	07	08	28	68
6.	−06	07	55	−02	27	23	17	21	26	04	03	13	00	60
7.	20	02	08	17	−05	25	09	18	42	−15	17	21	29	54
8.	29	14	02	−01	−05	25	−04	16	42	00	50	03	−15	65
9.	21	−01	09	11	−04	17	11	08	33	23	47	05	−03	50
10.	30	09	07	17	−03	29	04	19	32	02	59	00	−12	72
11.	42	21	−05	05	17	08	56	14	14	−01	−01	19	13	67
12.	42	01	03	05	12	−05	50	11	30	06	06	09	12	58
13.	33	17	−14	27	−08	−08	42	41	06	−04	04	15	−05	62
14.	30	18	−07	28	−18	01	36	27	−06	−09	24	−05	01	51
15.	53	40	−07	20	04	−02	33	05	20	00	−03	02	14	66
16.	03	23	16	16	04	20	22	11	09	05	14	10	20	29
17.	22	16	20	20	20	53	17	15	14	09	03	−12	03	57
18.	25	23	19	00	22	35	23	16	07	09	−07	03	02	42
19.	−06	25	00	14	25	48	−01	05	00	38	15	15	−10	58
20.	−33	05	16	24	24	35	−11	−04	12	−03	−01	−01	−20	44
21.	00	24	26	08	27	30	25	25	00	07	10	11	12	46
22.	13	20	09	19	20	21	15	04	44	28	14	11	20	55
23.	14	16	33	40	08	09	04	16	24	07	12	−07	−01	44
24.	17	05	08	31	52	20	04	00	21	23	08	10	−15	58
25.	−02	13	43	−09	50	10	17	22	19	01	27	01	04	66
26.	11	15	09	06	43	13	18	35	26	−04	19	−10	10	53
27.	−06	17	36	09	37	04	22	20	−12	28	23	17	01	57
28.	24	11	12	07	59	03	04	22	04	−07	19	04	04	53
29.	13	01	26	25	21	09	−16	09	06	02	−10	62	18	66
30.	45	05	23	10	02	−09	−06	00	12	16	−08	47	19	58
31.	07	30	12	19	51	14	−07	10	08	18	22	21	−12	59
32.	−13	15	26	20	−03	−02	26	15	−03	20	−08	−06	−13	31
33.	16	−15	18	−03	28	21	29	08	30	−21	01	11	−03	44
34.	30	51	01	19	07	−08	38	13	21	01	00	08	26	68
35.	09	03	17	56	05	−07	15	−01	17	01	−01	03	−03	41
36.	32	08	25	40	29	21	14	12	13	08	18	−09	17	59
37.	−02	18	−06	19	26	21	15	05	−01	42	29	−07	08	48
38.	57	20	02	−06	−07	02	28	07	22	−02	06	01	15	53
39.	04	−01	43	08	06	01	−02	−07	−01	−05	04	01	03	21
40.	22	53	00	−03	−02	−06	17	09	11	−02	−01	−15	02	41
41.	23	31	06	05	06	12	40	03	03	07	02	06	08	35

† Decimal points have been omitted.

Correlations. Coefficients of reliability were computed for the entire sample. Most of the estimates were based on correlations between separately-timed, matched forms of the tests, by means of the Spearman-Brown formula. The estimates are shown in Table I. Frequency distributions were made for all tests; none differed markedly from normal. Pearson product-moment correlations were computed between all variables on a digital computer.

C. *The Factor Analysis.* Thirteen factors were extracted by means of Thurstone's complete centroid method. The distribution of residual values obtained after the extraction of the twelfth factor contained no term greater than ± ·07. The unrotated factor matrix is shown in Table II. First approximation positions were obtained for the reference axes by placing each of nine axes near a single marker test in an orthogonal frame. This was done analytically by appropriate transformation matrix. From this position two rotation solutions were obtained by the Zimmerman graphic method [17]. Solution A was obtained by aiming toward simple structure and positive manifold, additional consideration being given to meaningfulness of the solution in terms of the nature of the tests. Solution B was obtained by using coded numbers for tests and reference axes, and considering only the improvement of simple structure and positive manifold. Only solution A was adequate for interpretation purposes. Its rotated factor matrix appears in Table III.

V. INTERPRETATION OF THE FACTORS

The interpretation of the factors rests mainly on loadings of ·30 or over. For each factor we give a list of tests with loadings of ·30 or over, with their names and numbers. Figures in parentheses are the loadings of ·30 or over for the tests on other factors.

Factor A. Verbal Comprehension (VC)

38.	Verbal Comprehension	·57	
15.	Inventive Verbal Relations	·53	(GR ·40, CC ·33)
3.	Word Listing III	·52	(WF ·57)
30.	Word Synthesis (rating)	·45	(SW ·47)
11.	Vocabulary Completion	·42	(CC ·56)
12.	Inventive Opposites	·42	(CC ·50, AF ·30)
4.	Word Listing II	·39	(WF ·61)
13.	Associations III	·33	(CC ·42, WF ·41)
36.	Plot Titles (cleverness)	·32	(O ·40)
1.	Rhymes	·30	(WF ·58)
10.	Controlled Associations—IIIe	·30	(AF ·32, Sc ·59)
14.	Associations IV	·30	(CC ·36)
34.	Verbal Relations Naming	·30	(GR ·51, CC ·38)
20.	Plot Titles (low quality)	·33	(IF ·35)

The most obvious characteristic of this factor is that many tests correlate with it. The reason lies in the nature of the battery which was devoted almost exclusively to verbal tests. An interesting negative loading is that for Plot Titles (low quality). It correlates negatively with the first eight tests in the list above, with coefficients between $-\cdot 13$ and $-\cdot 24$. It also had a negative loading in the cross-verification study (9) which reflected negative correlations between test variables. It may be that examinees who are high on *verbal comprehension* disdain to give prosaic titles for stories, or that the scorers of the titles are influenced by the superior *verbal comprehension* they find in some of the titles, rating such responses clever rather than of low quality. A more fundamental hypothesis would be that the more words we know, the harder it is to get out the most ready ones—a hypothesis supported by the work of Bousfield *et al.* [2].

Factor B. General Reasoning (GR)

40.	Ship Destination	·53	
34.	Verbal Relations Naming	·51	(VC ·30, CC ·38)
15.	Inventive Verbal Relations	·40	(VC ·53, CC ·33)
41.	Word-Group Naming	·31	(CC ·40)
31.	Word Arrangement	·30	(EF ·51)

Ship Destination was included in the battery because of the possibility that some of the fluency tasks performed with high restriction would contain a substantial proportion of reasoning variance. This has some basis in a study of reasoning tests by McNemar [10]. She found that differentiation between good and poor reasoners (separated by tests of *logical reasoning* and *general reasoning*) increases with the restriction placed upon the response. Tests 34, 15 and 41 are the most highly restricted tasks in the battery, each containing items with only one correct answer. The preferred definition for this factor has been the ability to comprehend or to structure problems in preparation for solving them. It would thus seem that an important condition for measuring *general reasoning* is high restriction in the response.

Factor C. Writing Speed (WS)

6.	Word Listing O	·55	
39.	Marking Speed	·43	
25.	Two-Word Combinations	·43	(EF ·50)
27.	Four-Word Combinations	·36	(EF ·37)
23.	Simile Completion	·33	(O ·40)

This is called *motor speed* by Bechtoldt [1] and is identifiable also with Rimoldi's [11] *speed at drawing with the hands*. Test 39 was the leading test in Bechtoldt's study. Here Word Listing O, in which the examinee lists words of any kind, appears as a strong measure of the factor; so are tests 25 and 27, which involve little restriction in producing combinations of words. The common feature of most fluency tests in this list is that little time is needed for thinking how to meet restrictions and more time devoted to writing.

Factor D. Originality (O)

35.	Cartoons	·56	
36.	Plot Titles (cleverness)	·40	(VC ·32)
23.	Simile Completion	·40	(WS ·33)
24.	Simile Interpretations	·31	(EF ·52)

Tests 35 and 36 have identified this factor in other studies. It is defined as the ability to produce remotely associated, clever, or uncommon responses. Simile Completion was included as a marker test for *expressional fluency*, but did not appear loaded on that factor. Simile Interpretations requires originality in thinking, since most of the objects in the test for which common attributes are to be expressed are only remotely associated.

Factor E. Expressional Fluency (EF)

28.	Four-Word Combinations FL	·59	
24.	Simile Interpretations	·52	(O ·31)
31.	Word Arrangement	·51	(GR ·30)
25.	Two-Word Combinations	·50	(WS ·43)
26.	Two-Word Combinations FL	·43	(WF ·35)
27.	Four-Word Combinations	·37	(WS ·36)

This factor obviously involves a facility in, or aptness for, producing connected discourse. Every test in the list involves production of word combinations. There is hardly any indication that the factor is concerned with high-quality responses. Tests 28, 31 and 26 suggest that producing connected discourse under restrictive conditions brings out the factor well. The advantage of test 28 over 27 as a measure can be explained by the fact that the artificial restriction placed on test 28 (requiring each word to begin with a particular letter) resulted in enough restriction in the flow of ideas, so that *writing speed* was not an important component. Perhaps relatively less writing reduced *writing speed* variance. This relation, in terms of presence and absence of *writing speed* variance, can be seen in comparing tests 25 and 26. Simile Interpretations (the simile-type test with greatest opportunity for combining words) appears one of the three good and relatively pure tests for this factor.

Factor F. Ideational Fluency (IF)

17.	Thing Listing II	·53	
19.	Brick Uses (fluency)	·48	(Sb ·38)
18.	Thing Listing I	·35	
20.	Plot Titles (low quality)	·35	(VC– ·33)
21.	Descriptive Completion	·30	

This factor is defined as the ability to evoke a large number of ideas in a situation in which there is little restriction. Brick Uses and Plot Titles are common reference tests. Thing Listing II is heavily loaded on it, a fact consistent with the definition. These tests place little or no premium on quality. From previous studies there remained some doubt whether Thing Listing tests measure *ideational* or *associational fluency*. Here there is no such confusion. Two of the three Thing Listing tests are loaded on this factor and not on *associational fluency*. The third Thing Listing test, which involves the highest restriction in the set, has no significant loadings on any factor, its communality being only ·28.

Factor G. Eduction of Conceptual Correlates (CC)

11.	Vocabulary Completion	·56	(VC ·42)
12.	Inventive Opposites	·50	(VC ·42, AF ·30)
13.	Associations III	·42	(VC ·33, WF ·41)
41.	Word-Group Naming	·40	(GR ·31)
34.	Verbal Relations Naming	·38	(VC ·30, GR ·51)
14.	Associations IV	·36	(VC ·30)
15.	Inventive Verbal Relations	·33	(VC ·53, GR ·40)

Tests 11, 12, 13 and 14 all require production of a word that fits a stated or implied relation. Although these tests were expected to be loaded on the factor *associational fluency*, they are not; nor are any tests loaded on *associational fluency* that are so highly restricted as to require uniquely determined responses. The only exception in this differentiation between tests of *eduction of conceptual correlates* and *associational fluency* is Inventive Opposites, which bridges the gap. It represents the characteristics of both factors, since, as with the tests listed above, it provides for near-unique responses by specifying the first letter of each antonym required, and also asks for two responses for each item. It thus furnishes the multiple-production characteristic of *associational fluency*. The separation of Vocabulary Completion and Inventive Opposites from multiple-production tests of *associational fluency* such as Controlled Associations was a genuine surprise when it occurred, since the two former tests have sometimes had loadings on *associational fluency*, and Inventive Opposites was selected as the marker for this factor. However, the correlation between all the Controlled Association tests and the two leading tests on this factor are low—a result which suggests a difference in function and provides a clear separation between the two factors.

Factor *G* can be defined as the ability to produce a response to fit a given or implied conceptual relation. As with the tests for *eduction of correlates*, the response restriction is so high that ordinarily only one response is correct for any item. The fact that many of the tests in the above list have loadings on *verbal comprehension* might suggest a correlation between these two factors. Three of the six *verbal comprehension* loadings are barely above the threshold for significance. Nevertheless, the factors are not well separated, and there may be a good reason. The best measures of *verbal comprehension* have been multiple-choice word-synonym tests. Most of the tests above involve finding the exact word to fit the given relation, usually a relation which involves meaning. One set emphasizes *selecting* a word meeting a requirement of meaning, the other necessitates *producing* it. Many individuals who can produce words fitting exact requirements of meaning would also be expected to excel in selecting words similar in meaning. The correlation between the factors may therefore be genuine; but it would appear from this study that they are separate.

Factor H. Word Fluency (WF)

2.	Suffixes	·71	
5.	Word Listing I	·62	
4.	Word Listing II	·61	(VC ·39)
1.	Rhymes	·58	(VC ·30)
3.	Word Listing III	·57	(VC ·52)
13.	Associations III	·41	(VC ·33, CC ·42)
26.	Two-Word Combinations FL	·35	(EF ·43)

All the tests in the list above represent production of words which meet structural rather than meaningful requirements, with the exception of Associations III, which has a relatively low loading. In addition, there is a certain amount of restriction in the tasks; but the loadings do not seem to be much affected by the variation of restriction imposed by Word Listing I, II and III. Within this range of variation, however, the test with the greatest restriction (Word Listing III) has the lowest loading, though the differences are slight. This condition also seems to increase the importance of *verbal comprehension.* The condition of *no* restriction (in Word Listing 0) gave an insignificant loading (·21) on *word fluency.* The inference is that a low to moderate level of restriction is optimal for tests of this factor: no restriction and high restriction are unfavourable.

Factor I. Associational Fluency (AF)

22.	Simile Insertions	·44	
7.	Controlled Associations IIIa	·42	
8.	Controlled Associations IIIbc	·42	(Sc ·50)
9.	Controlled Associations IIId	·33	(Sc ·47)
10.	Controlled Associations IIIe	·32	(VC ·30, Sc· 59)
33.	Alternate Headlines (word)	·30	
12.	Inventive Opposites	·30	(VC ·42, CC ·50)

Associational fluency has been discussed in relation to factor G (*eduction of conceptual correlates*). In contrast to that factor, *associational fluency* involves a multiple-association process. The relation that the words produced are required to satisfy is restricted as to meaning, but is not so far that only one or two responses will suffice. In the group of tests above the tasks that permit the greater number of responses are those with highest loadings. Calling up a number of words all similar in meaning or all fulfilling certain meaning specifications seems to be the essence of tests for *associational fluency*. Possibly that the factors identified as *associational fluency* in analyses by Thurstone [14], Zimmerman [16], Fruchter [6], and by Guilford, Kettner, and Christensen [8] were composites of two factors—*associational fluency* and *eduction of conceptual correlates*.

Factor J. Specific (Brick Uses) (Sb)

37.	Brick Uses (flexibility)	·42
19.	Brick Uses (fluency)	·38 (IF ·48)

The two scores from the Brick Uses test show a certain amount of spurious correlation, because the same responses are scored in two ways. This relation is taken care of by this factor. Brick Uses (flexibility) was put in the analysis to bring out a factor of *spontaneous flexibility*; but no such factor emerged.

Factor K. Specific (Controlled Associations) (Sc)

10.	Controlled Associations IIIe	·59 (VC ·30, AF ·32)
8.	Controlled Associations IIIbc	·50 (AF ·42)
9.	Controlled Associations IIId	·47 (AF ·33)

Tests 8, 9 and 10 correlate quite highly; and this factor reflects the source of portions of those correlations. These tests are also correlated with the factor of *associational fluency*, along with the marker Controlled Associations, test 7. Why test 7 does not share this specific variance is not clear; but this result seems reasonable when one considers the intercorrelations of the four test forms.

Factor L. Specific (Word Synthesis) (Sw)

29.	Word Synthesis (words)	·62
30.	Word Synthesis (rating)	·47 (VC ·45)

Factor M. Residual (R)

VI. Discussion

A. *Relation of the Factors obtained to the Differences Hypothesized.* In discussing the hypotheses and the relevant results two lists of factor loadings will be considered. List A gives factor loadings selected from Table III; list A′ gives the corresponding loadings corrected for unreliability. Correction for unreliability was preferred to correction for uniqueness, because reliability is independent of the collection of tests with which it happens to be analysed, whereas communality is not. Since each test was designed to stand for some hypothesized psychological function, and since tests differ in reliability, it is important when comparing factor loadings from different forms to allow for errors of measurement. As it turned out, the corrected loadings are in essentially the same rank orders as the uncorrected, and do not materially alter conclusions.

In comparing factor loadings, there is no agreed procedure for testing statistical significance. We can only note differences between factor loadings in relation to experimental variations, having more confidence when those differences are large and when there are trends, whether in line with predictions or not. We shall consider the special hypotheses in turn, considering the loadings of tests designed to test those hypotheses in connection with certain factors.

1. *Variation in Temporal Period of Production* (Controlled Association test forms)

	Loadings					
	A			A′		
	IF	AF	O	IF	AF	O
7. (First half-minute)	·25	·42	·17	·27	·45	·18
8. (Next 1·5 minutes)	·25	·42	−·01	·31	·52	−·01
9. (Third minute)	·17	·33	·11	·24	·47	·16
10. (Next two minutes)	·29	·32	·17	·34	·38	·20

The hypothesis was that the earlier the period of production the more IF variance a test should have; the later the period the more O variance; and for the intermediate periods more AF variance.

For IF there is no consistent trend in either temporal direction. The means of scores from tests 7 through 10 indicate increasing restriction, in terms of available supply remaining. The mean rate per minute for test 7 was 22·5 responses and the mean rates for tests 8, 9, and 10 were 7·5, 5·2 and 4·3. The failure of variance in IF to decrease systematically with time in these tests indicates that IF variance is not related to this kind of restriction in this kind of test.

The loadings on factor O are trivial, but such trend as there is, is for higher loadings at both the earliest and latest periods. The "trend" probably is insignificant and hence means nothing. But it can be inferred that the later, and hence more difficult, associations given in a Controlled Association test, and presumably more remotely associated, do not indicate individual differences in *originality* as is true in tests of different kinds.

There is just a trace of the expected relationship between the time of the responses during work on Controlled Associations and variance in AF, but we have to find it in the corrected factor loadings. The moderate time periods have the larger loadings, but only the drop after the third minute seems worthy of note. We can rest assured that Controlled Associations does as well as appears possible in measuring AF, at least during the first three minutes of work.

2. *Variations in Restriction of Class* (Thing Listing tests)

Test	AF	IF	AF	IF
16. (High restriction)	·09	·20	·15	·33
17. (Moderate restriction)	·14	·53	·19	·72
18. (Low restriction)	·07	·35	·11	·55

The degrees of restriction can be controlled in thing-listing tests in two ways. One is by the number of attributes specified for the class within which the things belong. Tests 16 and 17 had three attributes specified, e.g., things white, soft, and edible, whereas test 18 had two attributes. The other control, with number of attributes constant, is to find classes with varying numbers of potential responses. The success of our controls is indicated indirectly by the mean numbers of responses produced per minute. For the three tests these means were 1·4, 4·6, and 8·2, respectively.

The hypothesis was that the greater the restriction of the class the greater the AF variance and the less the IF variance. The prediction regarding AF variance failed; the AF variances were all near zero. The prediction of a monotonic trend for variance in IF also failed; the indication is that a moderate degree of restriction is optimal in a test for that factor. This suggests that, in addition to the variable of restriction vs. large potential supply of responses, there may be another variable working in the opposite direction. This might be the efficiency with which the examinee discriminates between class-members and non-members. A specified class is a kind of 'search model'. When potential is large, less attention need be paid to the model; when small and many exclusions are necessary the model becomes more important.

3. *Variations in Number of Structural Specifications in Isolated Words* (Word Listing tests, varying in number of specified letters)

Test	WF	IF	WF	IF	VC	WS	VC	WS
3. (Three)	·57	·00	·65	·00	·52	·01	·59	·01
4. (Two)	·61	–·04	·73	–·05	·39	–·06	·47	–·07
5. (One)	·62	·04	·72	·05	·21	·12	·24	·14
6. (None)	·21	·23	·22	·25	–·06	·55	–·06	·59

The hypothesis was that the greater the number of specified letters in a response, the higher would be the loading on WF and the lower the loading on IF. For WF there is no monotonic trend accompanying the variation. There is a marked difference in the predicted direction between the three tests involving one or more specified letters and the test involving none. Thus, at least one specified letter is essential for measuring WF. There is a slight indication that the optimal test for WF has one or two specified letters (or one syllable). As was expected, none of these tests was substantially related to IF. The loadings on this factor are essentially zero, except for test 6. With no restriction there is a trace of IF variance.

It is of interest to note variances in the non-fluency factors of *verbal comprehension* and *writing speed* in these WF tests. Neither was involved in our predictions. With factor VC, there is a progressive, positive trend towards higher loadings as a function of increasing restriction; naturally as more restrictions are added, the person with a larger vocabulary will have an advantage.

With factor WS, the trend is in the opposite direction: as words can be evoked at a greater rate, the limitations of writing speed becomes a contributor to variance in scores.

4. *Variation in the number of Connected Words*

Test	EF	WF	AF	EF	WF	AF
6. Word Listing O ..	·27	·21	·26	·29	·22	·28
25. Two-word Combinations	·50	·22	·19	·55	·24	·21
27. Four-word Combinations .. .	·37	·20	–·12	·41	·23	–·13
5. Word Listing I ..	·14	·63	·25	·16	·72	·29
26. Two-word Combinations FL	·43	·35	·26	·50	·41	·30
28. Four-word Combinations FL	·59	·22	·04	·76	·28	·05

In each trio, the requirement is the production of one word, two connected words, and four connected words, respectively. This variation is assumed to be along a dimension of increasing amount of continuous discourse. The first trio requires no initial letters, the second trio does.

It was hypothesized that the more words used together, within this range, the higher should be the loading on EF and the lower the loadings on WF and AF. With respect to EF, the hypothesis is supported, but the trend is complete only when first letters are specified. The hypothesis for WF is not supported in the first trio but strikingly supported in the second. This reflects the fact that there is no letter requirement in the first case but there is in the second. Some letter requirement is essential for the factor WF; but a significant loading on WF appears only for the condition of smaller (two-word) combinations. The loadings for factor AF agree with the hypothesis. The lack of significant loadings on AF for any of the tests reflects the fact that there is little requirement in those tests related to word meaning.

5. *Variation in Structural Specifications in Connected Discourse* (Word-Combinations tests)

Test	IF	EF	WS	IF	EF	WS
25. (Two words) ..	·10	·50	·43	·11	·55	·47
26. (Two words, FL) ..	·13	·43	·02	·15	·50	·11
27. (Four words) ..	·04	·37	·36	·04	·41	·40
28. (Four words, FL) ..	·03	·59	·12	·04	·76	·15

Bureau of Ednl & Psyl. Research
(S. C. E. R. T.) J851
Date ....

Here the variation is between tests within each pair, first and second. It was hypothesized that with no structural specifications there should be relatively more IF variance and with the first letter specified, more EF. Addition of the restriction in these tasks did not reduce IF variance; there was none any way. As to the EF variance, adding the restriction increases this component when there are four words to be combined, but slightly decreases it, or makes no difference, when there are two words. Again we find a variation more effective in affecting variances when other conditions (more words in the combination) are favourable for emphasizing the factor. The effect of the restriction on factor IF may interact with the effect upon another factor, namely, WS. WS is relatively more important in the two-word condition and without the first-letter restriction. It is a common observation, that an increase in variance in one factor is likely to be accompanied by a decrease in some other, as if communalities tended to remain relatively constant. It might be, therefore, that some *ideational fluency* variance could appear in the less restrictive form if some of these tests were responses given orally instead of being written.

6a. *Variation in the Process of Completing Similes* (Nouns vs. adjectives)

Test	AF	IF	EF	AF	IF	EF
23. Simile Completion (nouns)	·24	·09	·08	·28	·10	·09
22. Simile Insertions (adjectives)	·44	·21	·20	·52	·25	·24

The hypothesis was that producing *objects* in a given context should favour greater EF variance, whereas producing *attributes* should favour variances in AF and IF. For AF the hypothesis is strongly supported. The rate of production is 5·3 per minute for adjectives vs. 2·7 per minute for nouns, which should favour factor IF in test 22, according to the hypothesis. The results are in this direction, but in both cases IF variance is trivial. The results for EF contradict the hypothesis, with EF variance low in both tests. Instead of test 23 having a significant loading in EF, as in previous studies, it differs from test 22 by having a significant loading on *originality*. This suggests that some factors previously called EF could have been a confounding of EF and *originality*.

6b. *Variation in the Process of Completing Similes* (One word vs. several)

Test	AF	IF	EF	AF	IF	EF
22. Simile Insertions ..	·44	·21	·20	·52	·25	·24
24. Simile Interpretations	·21	·20	·52	·24	·23	·60

The hypothesis was that producing single-word attributes (as in test 22) should involve relatively more AF and IF variance, and producing multi-word expressions of attributes (as in test 24) should involve more EF variance. It is supported for AF and EF, but not for IF. The comparison brings out one important difference between factors AF and EF—the production of one-word responses vs. connected multi-word responses. The variation seems quite irrelevant to factor IF. This fact is also demonstrated in tests like Thing Listing (one-word responses) vs. Plot Titles (multi-word responses).

7. *Variation in Quality of Production* (vs. quantity)

Test	EF	IF	O	EF	IF	O
20. Plot Titles (low)	·24	·35	·24	·26	·38	·26
36. Plot Titles (high)	·29	·21	·40	·34	·24	·46
29. Word Synthesis (words)	·21	·09	·25	·25	·11	·29
30. Word Synthesis (rating)	·02	−·09	·10	·03	−·12	·13

The first variable of each pair is the low-quality or the quantity score; the second is the quality score, that for test 36 indicating cleverness and that for test 30 indicating good integration or coherence. It was hypothesized that quantity or low-quality scores would correlate more with IF, and quality scores more with EF and O. If both pair of tests are considered, there is no indication that EF depends upon quality of responses of the kinds considered here. If we consider only the results from Plot Titles, the expectation was realized with regard to IF and O, as usual. The loadings for the second pair of tests are generally very low, and offer no indication except the curious reversal of expectation with respect to *originality*. Here the low-quality score appears to have a slight edge. Possibly high quality in the form of rated coherence does not have the same bearing as rated cleverness of responses; perhaps clever responses are even less coherent than non-clever. On the whole, the quality-quantity variation seems irrelevant for factor EF, but not for the other two, when expressed in terms of rated cleverness.

8. *Variation in the Number of Connected Sentences*

Test	EF	IF	EF	IF
29. Word Synthesis (words)	·21	·09	·25	·11
31. Word Arrangement	·51	·14	·61	·17

The production of related sentences vs. unrelated sentences is the contrasting condition in tests 31 and 29. The former allows ten words to be used in any way in a sequence of sentences; the latter requires that the same four words be used in every sentence. The variation in restriction gave rise to the hypothesis that IF would be more strongly related to test 29; and it was expected that EF would be more related to test 31. This hypothesis is similar to that in variation 5, except that there the restriction was symbolic, whereas here it is meaningful or semantic. The hypothesis is supported for EF but not for IF, which has almost no relation to either test. It appears that a good test of EF should emphasize the act of sentence forming and not the total quantity of words that an examinee puts out in connected discourse. Again, restriction of just any kind seems irrelevant to IF.

9. *Rearrangement of Words vs. Substitution of Words* (Alternate Headlines test)

Test	AF	EF	AF	EF
32. (Rearrangement)	−·03	−·03	−·04	−·04
33. (Substitution)	·30	·28	·43	·40

It was predicted that the arrangement score would correlate more with EF and the substitution score more with AF. As it turned out, both had zero relation to the rearrangement score and marginally significant correlations with the substitution score. The lack of relation of the rearrangement score to EF is surprising, in view of the strong loading for EF in the Word Arrangement, The main difference in the two tests is that in the latter the same four words must be included, with other words added to make sentences, whereas in Alternate Headlines the same idea must be expressed in different ways, without necessarily using the same words. The difference in relation to EF suggests that the factor depends upon the examinee producing his own idea for a sentence as well as, or even in place of, organizing words in a sequence. But producing alternative ideas is the essence of factor IF, which might be expected to be a component of test 31 but not of 32. The relation of EF to finding sentence ideas vs. wording of given ideas remains a mystery.

10. *Synonymous vs. Non-Synonymous Associations* (Associations III and IV)

Test	AF	VC	CC	AF	VC	CC
13. (Synonymous) ..	·06	·33	·42	·07	·40	·51
14. (Non-Synonymous) ..	−·06	·30	·36	−·08	·38	·46

Here we have two highly restrictive associations tests, each calling for a unique response to every item, one requiring synonyms, the other non-synonyms. It was expected that test 13 would involve more VC variance, and that both tests would be significantly loaded on VC. Such were the findings, though the difference in VC variance was small. In both the associative-reaction format of the items leads one to expect some AF variance; but this was not the case. Instead, we find evidence that AF is a multiple-response affair. Neither test calls for multiple responses. They consequently have significant variances in factor CC, *eduction of conceptual correlates.*

B. *General Conclusion Regarding Tests of Hypotheses.* It will have become apparent that many of our hypotheses were given wrong slants, owing to the nature of previous findings in regard to fluency factors. Previous results had often found two or three fluency factors loaded significantly with the same test, either in the same or different analyses. This led us to believe that what made the difference in factorial content was not so much the type of test content as certain testing conditions which apply to tests of different content. Conspicuous among these conditions is some kind of restriction. The kind of restriction differs somewhat from one test to another; but several of the hypotheses led us to expect that minimal restriction favoured *ideational fluency* and maximal restriction favoured either *word fluency* or *expressional fluency*; *associational fluency* was expected where the restrictions were moderate.

As things turned out, every kind of fluency seems best measured by a test having a moderate degree of restriction. The differentiation in fluency factors is not associated with degree of restriction, but depends on the nature of the content and the operations to be performed—listing words, ideas, or sentences. Each requires a multiple-response task. Restriction to one unique answer is too much in every case. The fact that there must be some restriction means that the responding is done by calling up from memory storage information that satisfies some specified search model. Further work on fluency of production should keep these guiding principles in mind.

C. *Relation of Factors Obtained to Factors Previously Found.* Of the nine factors identified in this study, all have been identified before. Its value therefore lies in the new information concerning the nature of the factors, particularly the fluency factors, and of the best method of measuring them. Brief comparisons of the factors found with those previously reported may be of interest.

Word fluency is found to emphasize the production of words as letter combinations rather than words as meanings. Our results have made a sharper division between word-fluency tests and tests of other fluency factors. We found no tests loaded significantly on both *word fluency* and *associational fluency* or *ideational fluency*. In the past, there has been some question as to how much restriction there should be in the word-structure task for greatest word-fluency variance. Our study indicates that some restriction is necessary, but that, if the restriction is too great, a relatively high proportion of *verbal comprehension* variance becomes involved. The condition of no restriction emphasizes the factor of *writing speed.*

The definition of *ideational fluency* is not materially changed: tests like Thing Listing seem most characteristic. In previous studies there was some confusion with *associational fluency*, Thing Listing being sometimes loaded on both. It now seems certain that tests of *ideational fluency* specify meaningful classes. Some restriction in class size is necessary; classes of moderate size appear to be optimal. There is no need for high quality of responses; the number of low-quality responses continues to be the most nearly unique measure of this factor. The number of high quality responses brings in relatively more variance in *originality*.

The *associational fluency* found in some studies has been confounded with *eduction of conceptual correlates.* In this study we see that *associational fluency* is distinct from the correlates-eduction factor, emphasizing multiple-production in tasks involving finding related words; *eduction of conceptual correlates* involves the finding of a unique response. The former comes under the general heading of ' divergent production ', which applies to all the fluency factors; the latter under ' convergent production '.

Expressional fluency involves the production of connected discourse in the form of phrases or sentences. Even putting two words together in each item involves the factor, but more words are better; just how many would be optimal has still to be determined. High quality of responses is not an important attribute of scores for tests of this factor, as was sometimes formerly suggested. This study did not answer the question whether *expressional fluency* is a matter of organizing word sequences as structural or symbolic organizations, or a matter of organizing meanings, or perhaps both. Later studies should provide an answer.

REFERENCES

[1] BECHTOLDT, H. P. (1947). Factorial investigation of the perceptual speed factor. *Amer. Psychologist*, II, 304–5.

[2] BOUSFIELD, W. A. and SEDGEWICK, C. H. W. (1944). An analysis of sequences of restricted associative responses. *J. Gen. Psychol.*, XXX, 149–165.

[3] CARROLL, J. A. (1941). A factor analysis of verbal abilities. *Psychometrika*, **6**, 279–307.

[4] CHRISTENSEN, P. R., GUILFORD, J. P. and WILSON, R. C. (1957). Relations of creative responses to working time and instructions. *J. Exper. Psych.*, LIII, 82–88.

[5] FRENCH, J. W. (1951). The description of aptitude and achievement tests in terms of rotated factors. *Psychometric Monogr.*, No. 5. Chicago: University of Chicago Press.

[6] FRUCHTER, B. (1948). The nature of verbal fluency. *Educ. Psychol. Measmt.*, VIII, 33–47.

[7] GUILFORD, J. P. and CHRISTENSEN, P. R. (1956). A factor-analytic study of verbal fluency. *Rep. psychol. Lab.*, No. 17. Los Angeles: University of Southern California.

[8] GUILFORD, J. P., KETTNER, N. W. and CHRISTENSEN, P. R. (1959). The relation of certain thinking factors to training criteria in the U.S. Coast Guard Academy. *Educ. Psychol. Measmt.*, XIX, 381–394.

[9] KETTNER, N. W., GUILFORD, J. P. and CHRISTENSEN, P. R. (1959). A factor-analytic study across the domains of reasoning, creativity and evaluation. *Psychol. Monogr.*, No. 9.

[10] MCNEMAR, OLGA (1959). An attempt to differentiate between individuals with high and low reasoning ability. *Amer. J. Psychol.*, LXVIII, 20–36.

[11] RIMOLDI, H. J. A. (1951). Personal tempo. *J. abnorm. soc. Psychol.*, XLVI, 283–303.

[12] ROGERS, C. A. (1953). The structure of verbal fluency. *Brit. J. Psychol.*, XLIV, 368–380.

[13] TAYLOR, C. W. (1947). A factorial study of fluency in writing. *Psychometrika*, XII, 239–262.

[14] THURSTONE, L. L. (1938). Primary mental abilities. *Psychometric Monogr.*, No. 1.

[15] WILSON, R. C., GUILFORD, J. P., CHRISTENSEN, P. R. and LEWIS, D. J. (1954). A factor-analytic study of creative-thinking abilities. *Psychometrika*, XIX, 297–311.

[16] ZIMMERMAN, W. S. (1953). A revised orthogonal rotation solution for Thurstone's original primary abilities battery. *Psychometrika*, XVIII, 77–93.

[17] ZIMMERMAN, W. S. (1954). A simple graphical method for orthogonal rotation of axes. *Psychometrika*, XI, 51–55.

Vol. XVI
Part I

The British Journal of Statistical Psychology

May
1963

IN SEARCH OF AN EXPLANATION OF SOCIAL MOBILITY[1]

By MICHAEL YOUNG

Institute of Community Studies, Bethnal Green

and

JOHN GIBSON

Department of Genetics, Cambridge University

Studies in different countries, which have shown that rates of mobility are more alike than they were expected to be, have raised questions about the mechanisms underlying social mobility. The following paper puts forward the suggestion that social mobility belongs to a cybernetic mechanism whereby something like a 'steady state' is maintained in each occupational class by means of constant movement into and out of it. The quality discussed is 'intelligence'; and the argument stems from two main observations. The first is that the higher the occupational class, the higher is the measured intelligence. The second is that the children of parents towards the extreme show a 'regression' from that extreme towards the mean of the general population. If these two observations hold true over a period of time, it follows that there must be social mobility. In a third part of the paper it is argued that differential fertility between the classes gives rise to additional mobility. The last section discusses longer-term changes which may influence the amplitude of short-term oscillations between any two generations, and finally notes a certain parallelism with the laws of thermodynamics.

I. APPROACH TO THE SUBJECT

Sociological researchers have during the last decade made many measurements of rates of social mobility for different generations and in different countries. For comparisons between generations in one country the Glass report [1] (see also Sorokin [2]) is still the best source. When ten-year age-cohorts born from before 1890 to after 1920 were compared, very little difference was found in rates of mobility. Roughly the same proportion in the younger as in the older cohorts had occupational status higher than, or the same as, their fathers: 'in general the picture of rather high stability over time is confirmed'. More recently Lipset and Bendix [3] have, in a comprehensive review of the literature, compared different countries. They measure the rate of social mobility upwards by the proportion of the sons of manual workers who become non-manual, downwards by the proportions of the sons of non-manual workers who become manual. In urban areas rather similar proportions of sons move up in a number of countries—in France 35 per cent, in Germany 26 per cent to 30 per cent, in Switzerland 44 per cent, in Sweden 29 per cent, in Japan 33 per cent, and in the United States 31 per cent to 35 per cent. There is more discrepancy in the proportions moving down, varying from 13 per cent to 38 per cent. Studies in particular cities also show rather similar rates in Poona and in Tokyo, in Sao Paulo and Kansas City, in Aarhus and Indianapolis. A recent study of a London suburb, Woodford, produced the same results [5]. If, instead of a two-class system, a division is made into several more classes, the rate of mobility—say, into and out of the professions—would probably be less strikingly similar [6]. But at any rate the general conclusion of the work done so far is that rates of mobility, over time and space, are much more alike than they were expected to be.

---

[1] We should like to acknowledge the help received from Christopher Wallis with earlier drafts of this paper; from Sebastian and Mary Halliday and May Brenner with the small survey in Cambridge reported in the text; from the people who commented on drafts, including Dr. Sidney Brenner, Sir Cyril Burt, Hubert Child, E. Cooper-Willis, Professor Ernest Gruenberg, Dr. Frank Hahn, Dr. A. H. Halsey, Sir Julian Huxley, and Professor Thoday; and from the Joseph Rowntree Memorial Trust which met the costs of the Cambridge survey.

Now that some measurements have been made, however imperfect, the emphasis in the next phase of research should be on finding out what are the mechanisms underlying social mobility. But obviously some ideas will be needed beforehand about what these mechanisms might be. The purpose of this paper is to put forward one such suggestion, which, if borne out, would perhaps go some way towards explaining the similarity in mobility rates. We propose to test this suggestion in field research. We are both aware of the excesses of Social Darwinism in the 19th century, and concerned to prevent a repetition of them in the 20th; but we are also both keen to conduct hybrid research of interest to a biologist as well as a sociologist. Our hope is that other investigators in other places may be tempted into similar inter-disciplinary research.

Cybernetic Mechanism. The suggestion is, in brief, that social mobility is the variable element in what Cannon [7] called a homeostatic and Wiener [8] a cybernetic mechanism, whereby something like a ' steady state ' is maintained in each occupational class by means of constant movement into and out of it. The mechanism is in some ways similar to that envisaged in natural selection in the Malthusian scheme and in Ricardian price theory. We shall discuss how it might work in relation to the distribution, first of intelligence, and secondly of fertility, between occupational classes.

We are here defining ' intelligence ' rather narrowly as the qualities measured by the admittedly highly fallible instrument of intelligence tests, orthodox ones, not of the type which tests ' creativity ' [9]. This is not the only quality to take into account in any full investigation. The tenor of what Fisher said about what he called ' social promotion ' rather than ' social mobility ' is clearly right, even though the expression now sounds somewhat Victorian: " What is perhaps more important is that a number of qualities of the moral character, such as the desire to do well, fortitude and persistence in overcoming difficulties, the manliness of a good leader, enterprise and imagination, qualities which seem essential for the progress, and even for the stable organization of society, must, at least equally with intelligence, have led to social promotion " [10]. Such characteristics are important because they affect not only 'social promotion' but also the adaptability of a class as much as of a society. As Thoday has said, "We cannot assess just how much genetic diversity is needed in our complex and ever-changing social systems, but we can be sure that it is a good deal" [11]. Yet we have to start somewhere; and intelligence, although not the sole quality to consider, is at least an important one.

II. Mobility and Intelligence

Our argument stems from two main observations. The first is that the higher the occupational class, the higher the measured intelligence (I.Q. for short). The usual figures of this sort, for instance those collected by Floud, Halsey and Martin [12], are for children only. It has been shown again and again in many different countries that the children of (say) professional class parents score better than those of manual workers. There is less published information for adults: most intelligence tests are designed for children, not adults; but what there is, for instance, on the testing of recruits for the Armed Services [13], shows that there are the same sorts of differences for adults, although always more marked for the adults in a particular class than for the children. The figures in Table I illustrate this from a paper by Burt which marked an important advance [14].

Table I. Mean I.Q.s of Parent and Child According to Class of Parents

	Parent	*Child*
Higher Professional	139·7	120·8
Lower Professional	130·6	114·7
Clerical	115·9	107·8
Skilled	108·2	104·6
Semi-skilled	97·8	98·9
Unskilled	84·9	92·6
Average	100·0	100·0

Galton's Law. The second observation is that stated in Galton's 'law of regression' [15], namely, that the children of parents towards the extreme show a 'regression' from the extreme towards the mean of the general population [16]. Fathers (or mothers) with mean I.Q.s of about 84, or 140, have children with I.Q.s of about 92, or 120, respectively. This again is illustrated by Burt's table. Since the children from any particular class have a greater spread of intelligence than their parents, the shape of the distribution for the population in general is not progressively pinched, but remains roughly the same from generation to generation. Another way of putting it is to say that there is a correlation of less than unity between the I.Q.s of parents and children. From

Galton onwards many investigators have found correlations in the region of 0·5 between parent and child, and between sib and sib, not only for I.Q., but also for other graded characteristics, such as height and span of arms.

Why regression?[1] Is it due to genetic influences, to the effects of mating, or to environmental influences? We shall say a word about each. First, the genetic variable: (see Fisher [17] and Penrose [18]). The correlations of around 0·5 which have been found are about what would be expected theoretically if there were complete hereditary determination of I.Q.—but only on four assumptions, that several identical genes with additive effects are involved, that the gene pairs do not show dominance or recessiveness, that there is no assortative mating, and no environmental variations. The first two assumptions are difficult to test, though, as Burt and Howard [19] have shown, there is good evidence that the genetic element is mainly multifactorial. But the last two assumptions are unrealistic. Mating is assortative; the correlation between husbands' and wives' I.Q.s is about 0·5 (see Dahlberg [20]). The parent-child correlation should therefore be considerably above 0·5. It follows that the fourth assumption must also be unrealistic. The environments of parents when children and those of their own children are often similar, and this in part accounts for the similarity in intelligence; the fact that they are not identical accounts for much of the dissimilarity, and so tends to bring the correlation back to 0·5. Parents towards the extreme of the I.Q. distribution probably had exceptionally favourable or unfavourable environments and on chance grounds alone their environments are likely to be more exceptional than those of their children.

Necessity for movement. So we can say quite firmly that genetic, mating, *and* environmental influences on parent-child correlations are at work, even though in the present state of knowledge their respective effects may be difficult to sort out. Fortunately all that is necessary for our argument is the fact of regression. Given that, and given that the first observation we have referred to—the positive correlation between class and intelligence—is one that applies to different points of time, it follows that there must have been mobility upwards of more intelligent children to take the place of the less intelligent who move down from the higher classes: (for a review of the somewhat meagre evidence see Burt [21]). Had such mobility upwards and downwards not occurred, the mean I.Q.s for the parents in the upper classes would not be as high as appears in Table I, but would be at least as low as the means for the children, or indeed, if the regression had continued without compensation for a series of generations, they would by now have reverted to the general average; and the distribution of intelligence in each class would have become much more like the distribution in the general population than it actually is. A rough stability can only be maintained within each class by a continuous interchange between them: (cf. [22] and [23]).

---

[1] 'Regression' is partly due to the inevitable scatter of measurements. Even when the same person takes the same test twice, the correlation is less than unity.

In the absence of enquiry, we do not know what has been happening over time to the distribution of intelligence in the various classes. But we have made a small study of a kind not made before—an attempt to discover who, in I.Q. terms, actually moves up and down: this may serve to indicate the kind of material which could be gathered on a larger scale. A sample of 47 sons in their twenties was interviewed in 1962 in the town of Cambridge (students being excluded), and the 47 fathers in whatever part of England they were living. The results show that there was the same regression from fathers to sons as is found by all investigators, although less neat than that shown by Burt's figures (from a larger sample) in Table I. But the effect of regression was offset by ' exchange ' of sons between the classes in the way that our idea suggests should happen. Sons more intelligent than their fathers by and large moved up, and those less intelligent down; and, more important, if the distance of movement was measured on a 6-point class scale and related to the extent of the difference between the intelligence of fathers and sons, the greater the difference in intelligence, the greater the distance of movement. The consequence of this reshuffling was that the correlation between intelligence and class for non-manual workers as a whole was very nearly restored in the sons' generation to the level it had been in the fathers'. The correlation which had been 0·70 for the fathers came back to 0·68, as the result of mobility, for the sons. The Cambridge study therefore supports the suggestion we are making in this section of the paper.

III. MOBILITY AND FERTILITY

We will now briefly consider the effects of differential fertility between classes. Its relevance is obvious. If, for instance, intelligent adults in the upper classes do not by and large completely reproduce themselves, there will need to be movement upwards of intelligent children to maintain a stable correlation between intelligence and class. The first part of this proposition is probably not unrealistic (cf. [24]). It was at any rate accepted by the Royal Commission on Population: " Summing up what we may regard as reasonably well established in the evidence about the differential birth rate, it can be said that, of the social groups, those with the highest incomes, and among individual parents within each social group, the better educated and more intelligent have smaller families on the average than others " [25]. Glass [1] estimates how much mobility there would have to be to counter the effects of such differential fertility as was not fully offset by differential class mortality. Over the period with which he was concerned 300 non-manual fathers would have been replaced by only 252 adult sons, and 700 manual by 748 adult sons. To maintain stability, 5 per cent of sons would therefore have had to move up, and 5 per cent down, on this score alone.

Up till recent times it has been assumed that, if mobility was perhaps maintaining stability in the short run, it was only at the cost of disrupting it still further in the long run. Although not put in these terms, the idea was that the cybernetic mechanism was liable to a kind of increasing oscillation or 'hunting', with every correction creating a need for greater correction. This could have been the case on the ground advanced by Fisher: "That the economic situation in all grades of modern societies is such as favours the social promotion of the less fertile" [10]. For Fisher this was the reason why all civilizations have been eroded. For many others the same line of thought, buttressed by many studies which have shown a negative correlation between family size and intelligence, has suggested that, taking all classes together, intelligence must be declining [26].

Minnesota Study. This conclusion has, however, been called in question, both by comparisons of U.S. Servicemen in the two World Wars [27] and by comparisons over time of schoolchildren in Scotland [28]. These suggested no significant change in intelligence. More recently, Higgins, Reed, and Reed [29] in Minnesota have shown that much of the reasoning about intelligence decline has been questionable because it has not taken full account of childless people. They show that, although there is a negative correlation between intelligence of children and the number of their siblings, this does not result in any fall in intelligence because some of the most unintelligent children do not reproduce themselves. "The higher reproductive rate of those in the lower I.Q. groups who are parents is offset by the large proportion of their siblings who never marry or who fail to reproduce when married."

In any case the situation has been changed by the recovery in the birth-rate in the higher classes which has occurred since the Hitler war in the United States, and (to a lesser but still marked extent) in Britain and some countries of the Commonwealth and Western Europe. No-one yet knows the reasons for this change. All that seems certain is that biological considerations will have to be taken into account along with economic and sociological evidence. From the standpoint of biology, what has happened could be partly explained by two different arguments. The first is that the process has been in some ways similar to what happens when there is artificial directional selection in other species for a specific trait. Fertility is reduced for a time at the extreme high side of the distribution of the population until eventually a recovery is brought about by the restoration of genetic balance. The same thing could have happened with human beings precisely as the result of the stress given to the single trait of intelligence since the Industrial Revolution. As a result, mating has probably been more assortative for intelligence than before. This could have precipitated a decline in fertility of the more intelligent followed by a recovery. The second argument allows for a crucial difference between man and other species—that conception can be, and is, limited voluntarily. As the use of contraception has spread downwards from the upper classes, family size has depended more and more upon willingness to have children. But it is possible that this quality of willingness is to some extent genetic. In that event one would expect that birth rates would fall, to begin with, in any class when it first took to contraception on a large scale, but would recover in time as the proportion of 'willing' people in the population increased. As Darlington said: "Previously children had been born to parents merely in accordance with their ability to beget and bear them. Now they were born to parents in accordance with their willingness. Thus for the first time in evolution, parents who did not want children, or want so many of them, were selectively disfavoured. Conversely, a selection began which favoured specifically the property of wanting children" [30]. One merit of this argument is that its strength could be tested by enquiry.

The upshot of this discussion is that differential fertility may not give rise to as much mobility in the future as it has in the past. If data similar to the Minnesota data are found elsewhere, the lower fertility of people who are very low in I.Q., and probably in class too, may occasion downward mobility. At the other end of the distribution, the recovery of fertility amongst people high in I.Q. and class may be reducing the amount of upward mobility. But even though differential fertility is not the prompter it was, it is obviously important enough to be taken into consideration in any future enquiry.

IV. THE DYNAMICS OF MOBILITY

The idea we have discussed would help to explain not only mobility in general but the somewhat surprising similarities between countries mentioned at the beginning of this paper. If the same cybernetic mechanism is at work in all societies, that would help to account for the likeness in rates. But the discussion has been cursory. Not only have we concentrated solely on one of the characters with a bearing on mobility, but we have not been by any means exhaustive about that. We have not, for example, considered the influence of the variance of intelligence within each occupational class: the greater the variance and the more the overlapping between classes, the more one would expect mobility to arise between them. Nor have we considered how the standards by which 'intelligence' is judged change from one generation to another: again, the more change there is in the criteria, the greater should be the mobility. Nor have we considered the influence of the change from one generation to another in the size of classes: if the classes at the top expand, mobility upwards must be increased.

We have for the sake of exposition presented the idea in terms of a simple homeostatic mechanism. This may be acceptable as far as it goes: there may be a tendency in the short run for return to a 'steady state' from generation to generation. But underlying the movement from one generation to another are almost certainly secular changes—the long-term changes which set the context for the short-term oscillation between any two generations. To be more specific, what we imagine is that before a society embarks on the process of industrialization, marked above all by the application of Weberian rationality [31] or reasoning power to social affairs, the distribution of intelligence in any one class is not so dissimilar from that in the population at large. But as the process of industrialization and the application of science gets under way, intelligence is elevated beyond all other human characters and its distribution between the classes is pulled apart. The ‘ upper classes ’ gain more than their proportionate share of intelligence; and with every increase, the rate of mobility required to maintain that share goes up. The ‘ equilibrium ’ or ‘ steady state ’ is, in other words, likely to be changing from one generation to another. But a limit or critical point may be reached, at which no further increase in mobility can be secured. This brings us back to the mobility studies already mentioned. Lipset and Bendix suggest that most of the industrial societies for which data exist may have reached such a critical upper limit [4]. When and if mobility does knock up against such a limit there would be no further rise in intelligence in the upper classes, and those classes might as a result lose some of their previous buoyancy. The rate of economic expansion might then become less. Only comparative studies in different countries in various stages of economic growth could bring definition to such a view.

Before finishing, we should say that there is an alternative way of describing what we have in mind which has the merit that the starting point is not the notion of a ‘ steady state ’. Substituting ‘ mental energy ’ for intelligence immediately suggests a parallelism with thermodynamics. There may, in other words, be some applications in society of the laws of thermodynamics. Energy more ordinarily conceived is, according to the First Law of Thermodynamics, conserved. There may be a similar tendency for mental energy to be conserved. We referred to this above when we mentioned ‘ genetic balance ’. And just as energy is, according to the Second Law of Thermodynamics, subject to diffusion or entropy, so too may there be a tendency for mental energy to be diffused over all occupational classes. But if there is anything in the argument put forward in this paper, the entropic tendency in society is countered by constant interchange of mental energy between the classes. This interchange allows the quantity of ‘ mental energy ’ or ‘ intelligence ’ stored in the form of knowledge to be augmented from one generation to another.

This paper is a prologue to enquiry. What is implied is the nature of the enquiry that is required. It would have to follow the small Cambridge survey (referred to previously) in gathering information by means of interview in two or more generations about I.Q.s and occupations of parents and offspring, and about their education. One of the purposes would be to show the influence of education. It would also have to differ from the type of enquiry made so far into social mobility. None of the studies has been able to deal explicitly with differential fertility. Questions have been put, in one generation alone, to samples of sons, the sons being questioned about the occupations of fathers. Since the comparison has been between father-son pairs, each father in each class has been shown as having been replaced by just one son. If the influence of fertility is to be investigated, it will be necessary to start with a sample of people old enough to have had children and proceed to trace as many as possible of the children they actually have. It is this kind of enquiry which will, we hope, yield systematic comparative information on regression, fertility and mobility. It is time sociologists did more than just measure mobility; it is time they tried to examine the forces which underly it.

REFERENCES

[1] GLASS, D. (ed.) (1954). *Social Mobility in Britain.* London: Routledge.
[2] SOROKIN, P. (1927). *Social Mobility.* New York: Harpers.
[3] HO, PING-TI (1962). *The Ladder of Success in Imperial China.* New York: Columbia University Press.
[4] LIPSET, S. M., and BENDIX, R. (1960). *Social Mobility in Industrial Society.* London: Heinemann.
[5] WILLMOTT, P., and YOUNG, M. (1960). *Family and Class in a London Suburb.* London: Routledge.
[6] MILLER, S. M. (1960) *Comparative Social Mobility.* Oxford: Blackwell.
[7] CANNON, W. B. (1932). *The Wisdom of the Body.* New York: North.
[8] WIENER, N. (1948). *Cybernetics.* New York: Wiley.
[9] GETZELS, J. W., and JACKSON, P. W. (1962). *Creativity and Intelligence.* New York: Wiley.
[10] FISHER, R. A. (1930). *The Genetical Theory of Natural Selection.* Oxford: Oxford University Press.
[11] THODAY, J. M. (1963). Causes and functions of genetic variety. *Eugenics Review,* LIII, 4–26.
[12] FLOUD, J. E. and HALSEY, A. H., and MARTIN, J. M. (1956). *Social Class and Educational Opportunity.* London: Heinemann.
[13] CONRAD, H. S. and JONES, H. E. (1940). *Yearbook Nat. Soc. Stud. Educ.,* XXXIX, 97–141.
[14] BURT, C. (1961). Intelligence and Social Mobility. *Brit. J. Statist. Psychol.,* XIV, 3–25.
[15] GALTON, F. (1889). *Natural Inheritance.* London: Macmillan.
[16] ANDERSON, C. A., BROWN, J. C. and BOWMAN, M. J. (1952). Intelligence and occupational mobility. *J. Pol. Econ.,* XL, 218–239.
[17] FISHER, R. A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Trans. Roy. Soc. Edin.* LII, 399–433.
[18] PENROSE, L. S. (1949). *The Biology of Mental Defect.* London: Sidgwick & Jackson.
[19] BURT, C. and HOWARD, M. (1956). The multifactorial theory of inheritance and its application to intelligence. *Brit. J. Statist. Psychol.,* IX, 95–131.
[20] DAHLBERG, G. (1948). *Advances in Genetics,* II.
[21] BURT, C. (1959). Class differences in general intelligence. *Brit. J. Statist. Psychol.,* XII, 15–34.
[22] CONWAY, J. (1958). The inheritance of intelligence and its social implications. *Brit. J. Statist. Psychol.,* XI, 171–190.
[23] HALSEY, A. H. (1959). Class differences in intelligence. *Brit. J. Statist. Psychol.,* XII, 1–4.
[24] BURT, C. (1952). *Intelligence and Fertility.* London: Cassell.
[25] *Report of Royal Commission on Population.* (1949). London: H.M. Stationery Office.
[26] ANASTASI, A. (1956). Intelligence and family size. *Psychol. Bull.,* LIII, 187–209.
[27] TUDDENHAM, R. D. (1948). Soldier intelligence in world wars I and II. *Amer. Psychol.,* III, 54–56.

[28] Scottish Council for Research in Education, XXX. (1949). *The Trend of Scottish Intelligence.* London: U.L.P.

[29] HIGGINS, J. V., REED, E. W. and REED, S. C. (1962). Intelligence and family size: a paradox resolved. *Eugenics Quarterly*, IX, 2.

[30] DARLINGTON, C. D. (1960). The future of man. *Heredity*, XV, 441–442.

[31] WEBER, M. (1947). *The Theory of Social and Economic Organisation.* Oxford: Oxford University Press.

Vol. XVI The British Journal of Statistical Psychology May
Part I 1963

A FACTORIAL STUDY OF VALUE JUDGEMENTS AMONG ADOLESCENT GIRLS

By H. J. HALLWORTH and G. WAITE
Education Department, University of Birmingham

The following paper reports an experimental study of what is, for brevity, termed ' evaluation '—i.e., the degree of approval or disapproval accorded by the subjects tested to certain familiar concepts, such as ' honesty ', ' kindness ', ' pleasure ', ' money ', ' religion ', and the like. The method used was a ' semantic differential ' test, given to 128 girls aged 14 to 15. Forty-eight concepts were presented for evaluation, and seven evaluative scales—each with seven grades or points—were used. The mean evaluative scores for each concept were correlated over all the persons tested. The correlation matrix was analysed by the method of principal components; this furnished a first component which was identified as a factor of ' general evaluation '. A rotation according to the Varimax criterion was then applied, and yielded five orthogonal factors which apportioned the variance more evenly. A psychological interpretation is suggested for each of these factors so obtained.

I. INTRODUCTION

What is technically called the ' semantic differential technique ' has been used in a variety of circumstances to measure the meaning assigned by different individuals to various concepts. Osgood [7] has shown that, when pairs of adjectives with opposite meanings are used to define seven-point scales and these scales are employed by subjects who rate a number of concepts, then the scales do not really function as discrete measures. Subjects apply such scales as if they were grouped into three main sets, the scales within any one set being regarded as more or less interchangeable. Each such set of scales may therefore be described as a ' dimension of meaning '.

In analysing subjects' ratings Osgood originally summed over both subjects and concepts, and correlated every scale with every other. The sets into which subjects could be said to have grouped the scales were then defined by obtaining centroid factors from the correlation matrix and rotating these orthogonally to simple structure. The three main factors so extracted were described as ' evaluative ', ' potency ', and ' activity ' factors. In cases where the scales have been applied to concepts which are people or policies, it has commonly been found that the last two factors coalesce into a single ' dynamism ' factor. The evaluative factor, however, has invariably appeared, and represents the subject's sense of value, i.e. his approval or disapproval of the various concepts.

In a recent review of the large number of studies which has now been made of ' meaning systems ', Osgood [6] indicates the apparent generality of such findings in different samples of people. As far as present studies go, sex, intelligence, language, and culture appear to have little bearing on semantic factor structure. Studies based on the 'semantic differential' are necessarily three-way correlational problems: they relate to subjects, concepts, and scales. There are, therefore, potentially three origins of variance in the factor structure; and it is possible to eliminate one of these in order to study, for example, the placing in semantic space of various concepts judged by a single subject and of a single concept judged by a number of subjects, as well as the differences in the semantic spaces of a number of subjects (or groups of subjects) who use the same scales to judge a number of concepts. Researches based on the 'semantic differential' have generally been designed to throw light on some such aspect of the structure of semantic space [6]. Sometimes they have been concerned with the underlying mediation theory of meaning [2].

It is, however, suggested that those scales which have been found to lie on the evaluative dimension may be used to obtain a measure of general attitude to a concept [7]. If a number of subjects use several such evaluative scales to judge a number of concepts, each subject may be allocated a mean evaluation for each concept by averaging his scores over all the scales. Upon correlating these evaluative scores for all concepts and factorizing the resultant matrix, we may obtain a description of the subjects' evaluative or attitudinal space. For many samples of subjects this 'space' is of considerable interest to psychologists. An investigation was therefore designed to obtain a description of this 'evaluative space' for adolescent girls.

II. Design of the Investigation

The Measuring Instrument. Seven scales were chosen which have consistently shown high loadings on Osgood's evaluative dimension. These were combined with the scales for activity and potency to make a set of thirteen seven-point scales, as follows (marked for 'HONESTY'):

E7	successful		×						unsuccessful	1
	still					×			moving	
	slow			×					fast	
E7	beautiful		×						ugly	1
E1	cruel						×		kind	7
	difficult			×					easy	
E7	clean			×					dirty	1
	manly			×					womanly	
E7	wise			×					foolish	1
E1	dishonest							×	honest	7
	hard				×				soft	
	light				×				heavy	
E7	good	×							bad	1

The evaluative scales are labelled ' E ', the low end of each scale being marked ' 1 ' and the high end ' 7 '. Details regarding the construction of this set of tests and the assessment of its reliability and validity have been reported elsewhere [9]. In the present investigation each subject was given a mean score for the evaluative scales alone. The inclusion of the non-evaluative scales served to divert her attention from the object of the investigation.

The scales were applied to the 48 concepts enumerated in Table I. Three concepts were included twice in different parts of the test, namely, ' books ', ' religion ', and ' teachers '. The assumption was that the duplicated concepts would have two comparable loadings on each factor, and so indicate the consistency of responses. The concentration required to answer the test ensured that few subjects noticed the repetitions. Certain pairs were assumed to be opposite, namely, honesty and lies, friend and enemy, success and failure, pleasure and pain, reward and punishment, kindness and hate. It was expected that, if the measure was working as intended, such pairs would appear with opposite loadings on any one factor.

Subjects. The individuals tested were 128 girls, aged between 14 and 15 years, in their fourth year of the secondary school: 86 were in a secondary modern school, and 42 in a nearby grammar school. Both schools were in a working-class suburb on the edge of a Midlands conurbation. The test was administered during one session of 60 minutes. The instructions were to the following effect, but explained and illustrated in fuller detail.

" The test is aimed at finding the meaning that different words will have. It contains a number of words, printed in capitals, such as ' HONESTY '. Underneath each is a list of 13 pairs of opposites (always the same): e.g., ' successful ' . . . ' unsuccessful ' " (see p. 38 above). " There are seven spaces between each pair. Put a cross in one of the squares to show where you think the meaning of the word in capitals would be. If very ' successful ', put the cross at the extreme left; if only moderately ' successful ', not so far to the left; if very ' unsuccessful ', to the extreme right; if neither, in the middle. And so on."

III. ANALYSIS OF RESULTS

Each of the 128 subjects received a mean evaluative score for each of the 48 concepts. By means of an electronic computer, inter-correlations were obtained for all 48 concepts. The correlation matrix was then analysed by the method of ' weighted summation ' or principal components. Five components were extracted, which together accounted for a total of 55·1 per cent of the variance. Loadings for the first component, which accounted for as much as 33·7 per cent of the variance, are shown in Table I.

Finally, the axes were rotated by the Varimax method[1] to yield the factors shown in Table II.

---

[1] The statistical analysis was performed on a Ferranti Mercury computer. The Autocode and machine code programmes for correlation matrices, principal components analysis, and rotation to the Varimax criterion were written by Hallworth. The 'normal' Varimax criterion and 'loose' limit are used in the rotation programme (Kaiser, 1959).

It may be objected that the criteria used in analytic methods of rotation are psychologically irrelevant and that some prediction should be made by specifying, for example, that certain loadings will be zero, or again that the Varimax criterion permits only orthogonal rotation and that the most appropriate factors would probably be oblique. To both objections the reply is much the same and relates to strictly practical grounds. The initial components are mathematically justifiable in that the maximum possible variance is extracted by each successive component. Mathematical respectability does not necessarily equate with psychological meaning: nor do the principal components necessarily make the best psychological sense. The Varimax criterion, which maximizes variance in the columns of the factor matrix to obtain an approximation to Thurstone's 'simple structure' has been found in practice to give results which are psychologically satisfactory [3]. Lawley's method of estimation with specific initial assumptions [5] might be preferred by some psychologists. Or a programme for oblique rotation, such as Oblimax [8] or Maxplane [1], might possibly be used in the hope of describing the data with greater economy. But no computer programmes have yet been written for any of these procedures in this country. It was therefore necessary to use such tools as were available; and the most efficient was the Varimax programme here employed. This method has the further merit of being the same as that used by Osgood for more recent work [6].

TABLE I. LOADINGS OF CONCEPTS ON FIRST PRINCIPAL COMPONENT

Concept	Factor loading	Concept	Factor loading
1. Honesty	·59	25. Rules	·67
2. Home	·61	26. Future	·57
3. Dress	·67	27. Men	·51
4. Kindness	·69	28. Friend	·69
5. School	·63	29. Girl	·65
6. Woman	·61	30. Adults	·78
7. Pleasure	·62	31. Cleanliness	·54
8. Books	·50	32. Youth	·37
9. Religion	·76	33. Neighbours	·46
10. Lies	−·63	34. Pain	−·48
11. Failure	−·53	35. Mother	·70
12. Freedom	·37	36. Mischief	−·39
13. Teachers	·61	37. Games	·57
14. Work	·59	38. Policeman	·69
15. Ambition	·57	39. Reward	·56
16. Play	·64	40. Study	·71
17. Enemy	−·43	41. Trust	·71
18. Boy	·39	42. Punishment	·11
19. Money	·46	43. Success	·76
20. Disappointment	−·40	44. Shame	−·26
21. Myself	·61	45. Religion	·75
22. Sister	·57	46. Fear	−·20
23. Hate	−·64	47. Teachers	·64
24. Father	·62	48. Books	·51

TABLE II. LOADINGS OF CONCEPTS ON FIVE VARIMAX FACTORS

Concept	I	II	III	IV	V
1. Honesty	·29	·21	·45	·49	·10
2. Home	·28	·56	·41	·10	−·16
3. Dress	·28	·66	·31	·15	−·07
4. Kindness	·38	·44	·27	·43	·01
5. School	·05	·25	·79	·16	·19
6. Woman	·21	·67	·20	−·02	·16
7. Pleasure	·15	·66	·08	·24	·20
8. Books	·00	·18	·35	·18	·56
9. Religion	·49	·40	·32	·24	·19
10. Lies	−·46	−·30	−·24	−·48	·11
11. Failure	−·30	−·26	−·09	−·61	·01
12. Freedom	−·05	·47	·03	·09	·29
13. Teachers	·02	·25	·82	·04	·27
14. Work	·37	·06	·45	−·04	·46
15. Ambition	·17	·64	−·01	·17	·24
16. Play	·35	·53	·13	·04	·29
17. Enemy	−·48	−·15	·12	−·45	−·02
18. Boy	−·01	·14	·23	·02	·64
19. Money	−·01	·42	·10	·34	·28
20. Disappointment	−·13	−·23	−·11	−·70	·15
21. Myself	·24	·50	·30	·04	·17
22. Sister	·20	·31	·31	·05	·26
23. Hate	−·67	−·07	−·15	−·44	−·09
24. Father	·52	·19	·33	·14	·10
25. Rules	·32	·22	·60	·06	·26
26. Future	·26	·52	·27	−·12	·16
27. Men	·19	·11	·40	−·08	·56
28. Friend	·58	·39	·19	·08	·15
29. Girl	·46	·51	·03	·02	·29
30. Adults	·40	·34	·47	·10	·40
31. Cleanliness	·77	·07	−·02	·21	·10
32. Youth	·26	·24	−·16	·05	·46
33. Neighbours	·35	·41	·15	−·37	·24
34. Pain	−·09	−·31	·04	−·67	−·24
35. Mother	·67	·30	·23	·16	·06
36. Mischief	−·20	·15	−·26	−·50	−·28
37. Games	·38	·48	·32	−·18	·01
38. Policeman	·70	·13	·42	·05	·09
39. Reward	·29	·16	·03	·34	·58
40. Study	·37	·21	·50	·16	·35
41. Trust	·65	·36	·14	·11	·16
42. Punishment	·17	−·05	·35	−·34	−·04
43. Success	·47	·36	·11	·38	·43
44. Shame	−·21	·14	·06	−·65	−·15
45. Religion	·59	·45	·22	·12	·14
46. Fear	−·01	·08	−·02	−·68	−·09
47. Teachers	·13	·24	·82	·02	·20
48. Books	·11	·23	·33	−·04	·57
Percentage of Total Variance	13·57	12·83	11·10	9·63	8·01

IV. Discussion

General Attitude Factor. Since all the scales used in this investigation were chosen because they had been found in previous work to fall consistently on the evaluative dimension, it was expected that a large evaluative or general attitude factor would appear. Five components were extracted. The first component, shown in Table I, accounts for one-third of the variance—more than four times as much as the second component, and approximately seven times asmuch as the third. The five concepts with greatest negative loadings on this component are as follows:

23. hate	−·64	34. pain	−·48
10. lies	−·63	17. enemy	−·43
11. failure	−·53		

These negative loadings are consistent with the suggested interpretation: each of the five concepts is one member of a pair among the six sets of ' opposites '. Of the other pair one concept has a low positive loading ('punishment'). All the concepts included twice have comparable loadings on the general factor on both occasions, as follows:

Books	8.	.50;	48.	·51
Religion	9.	·76;	45.	·75
Teachers	13.	·61;	47.	·64

The amount of variance extracted by the first component supports the hypothesis of a general evaluative factor, and the detailed loadings give some indication of the consistency with which the seven evaluative scales were used.

The remaining four components extract only 8·3, 5·0, 4·3 and 3·8 per cent of the variance respectively; and it is difficult to suggest a plausible psychological interpretation for any of them. Since the object of the investigation was to obtain a map of the evaluative space of the subjects it is, therefore, necessary to examine the Varimax factors. The loadings are shown in Table II.

Varimax Factor I: Virtues and Associated Personalities. The first Varimax factor accounts for 13·57 per cent of the variance. Concepts with high positive loadings are as follows:

31. cleanliness	·77	45. religion	·59	29. girl	.46
38. policeman	·70	28. friend	·58	30. adults	.40
35. mother	·67	24. father	·52	4. kindness	·38
41. trust	·65	43. success	·47		

Concepts with high negative loadings are as follows:

23. hate	−·67	17. enemy	−·48	10. lies	−.46

The factor is apparently related to cleanliness, godliness, uprightness, and the personalities associated with such virtues. Although the conclusion is not shown by the data, it is probable that policemen, parents, and friends are regarded as personifying such virtues. Item 21 (' myself ') has a low loading on this factor, whilst 'teachers', 'books' and 'school' have near-zero loadings. In some respects the factor might be regarded as a factor of security-insecurity; but further marker concepts would be required to define it more precisely.

Varimax Factor II: The Feminine Image. The second Varimax factor accounts for 12·83 per cent of the variance. Concepts with high loadings are:

3. dress	·66	16. play	·53	12. freedom	·47
6. woman	·67	26. future	·52	45. religion	·45
7. pleasure	·66	29. girl	·51	4. kindness	·44
15. ambition	·64	21. myself	·50	19. money	·42
2. home	·56	37. games	·48	33. neighbours	·41

This group of concepts is apparently related to ambitions and pleasures which are regarded as typically feminine and with which it would seem that these girls tend to identify themselves. The name suggested is the 'feminine image'.

Varimax Factor III: The Image of School and Authority. Varimax factor III accounts for 11·10 per cent of the variance. Items with high loadings are:

13. teachers	·82	30. adults	·47	38. policeman	·42
5. school	·79	1. honesty	·45	2. home	·41
25. rules	·60	14. work	·45	27. men	·40
40. study	·50				

The name suggested for this factor is the 'image of school and authority'. The persons studied have a tendency to make evaluations of teachers, school, and study comparable with their evaluation of rules and policeman. A comparison of factors I and III suggests that, with the latter factor, a policeman is considered in his role as an upholder of authority. Cleanliness and religion have low loadings. The moderate loading of home, and the somewhat lower loadings of kindness, woman, myself, sister, mother, father, future, and games suggest a more rounded and pleasant image of school than at first appears. However, the predominant indication is that school is viewed as a place of discipline and study; and this is confirmed by the moderately high loading of punishment.

Varimax Factor IV: Abstract Concepts. The fourth Varimax factor accounts for 9·63 per cent of the variance. The concepts with highest positive loadings are as follows:

1. honesty	·49	43. success	·38	19. money	·34
4. kindness	·43	39. reward	·34		

Those with negative loadings are as follows:

20. disappointment	−·70	11. failure	−·61	23. hate	−·44
46. fear	−·68	36. mischief	−·50	33. neighbours	−·37
34. pain	−·67	10. lies	−·48	42. punishment	−·34
44. shame	−·65	17. enemy	−·45		

These concepts are almost all abstract. Moreover, most of the abstract concepts included in the test have substantial loadings on this factor. The exceptions are cleanliness, trust, pleasure, ambition, future, and rules, which have high loadings on factors I, II, and III, and are therefore defined in terms of the other loadings on these factors. Freedom, has near-zero loadings, except on factors II and V.

The probable explanation of this grouping of abstract concepts into one factor is that the adolescent girls of this study were unable to differentiate between such concepts, and therefore made a global judgement of approval-disapproval on each. Possibly the factor may be regarded as a remnant of the general evaluative factor which has not been further sub-divided owing to the subjects' lack of verbal understanding. The positive loadings are only moderate in size; the larger loadings are all negative. This suggests that differentiation from the general evaluative factor takes place positively, and that it is primarily the ‘ disapproval ’ or ‘ avoidance ’ end of the factor which remains. The only other notable features are the positive loading of ‘ money ’ and the negative loading of ‘ neighbours ’. The latter, combined with the positive loading for the same concept in factor I, suggests an ambivalent attitude towards neighbours which does not exist towards other concepts.

Varimax Factor V: The Masculine Image. Varimax Factor V accounts for only 8·01 per cent of the variance. High loadings are as follows:

18. boy	·64	27. men	·56	43. success	·43
39. reward	·58	32. youth	·46	30. adults	·40
48. books	·57	14. work	·46	40. study	·35

This factor apparently marks the beginning of a ‘ masculine image ’. ‘ Books ’, ‘ work ’, and ‘ study ’ fall within this group; but there is no evidence as to the significance of ‘ reward ’ or ‘ success ’.

V. SUMMARY AND CONCLUSION

The foregoing inquiry has attempted to explore the ‘ evaluative space ’ of a sample of secondary school girls by modifying a technique already employed for other purposes by Osgood. The girls' approval or disapproval of a concept was measured by obtaining the projection of that concept onto the evaluative dimension of the semantic differential, the evaluative dimension being defined in terms of scales which in previous factor analyses have been found to have high loadings on this dimension. When attitude is measured in this way it might be assumed that we are using a uni-dimensional attitude scale. And on correlating subjects' scores and analysing the matrix into its principal components, the first component, accounting for one-third of the variance, appeared to justify such a conclusion, and was accordingly identified as a factor of general evaluation or attitude towards the concepts used in the present form of the semantic differential.

However, when the axes are rotated by the Varimax method, five orthogonal factors are obtained among which the variance is more evenly apportioned. These, as we have seen, can readily be given a psychological interpretation. As evaluated by this sample of girls, the concepts used in the test may be grouped according as they are part of (i) the virtues and their personification, (ii) an image of school and authority, (iii) feminine and masculine images, and (iv) abstract concepts. Virtues are associated with home and friends, an important virtue being cleanliness. Teachers and school are associated with study, work, rules, honesty and authority figures. The feminine image appears to be an image of the woman whom the girl would like to be; and woman is associated with dress, pleasure, ambition, home and freedom. The masculine image is poorly developed, its only important associations being with books, work, and reward.

If the subjects' evaluation of the concepts were strictly uni-dimensional, it would be impossible to obtain five orthogonal factors which apportion the extracted variance so evenly. This possibility appears to derive from two sources. First, the scales used are not pure evaluative scales: all such scales have been found in previous analyses to be contaminated by small loadings on other dimensions. Secondly, there are, as Osgood indicates [7], changes in the kind of evaluation as the frame of reference changes from concept to concept: in other words, there may be a rotation of the evaluative scales so that, when applied to some concepts, they tend to align with another dimension. In these terms each of the five Varimax attitude factors indicates the rotation of the evaluative scales in the same direction for all concepts having high loadings on that factor; and thus represents a different kind of evaluation common to certain sets of concepts. In the present analysis it is only possible to specify these evaluations by inference, after examining the particular concepts grouped for each factor. Further information could be obtained by analyses which included a wider range of concepts and scales from other dimensions.

The results obtained raise questions regarding the comparable factor patterns among adolescent boys, and the variations in pattern between individuals with wide differences in personality. The development of the necessary computer programmes has made such analyses practicable, and they are now being undertaken as extensions of the work described in this article.

REFERENCES

[1] Cattell, R. B., and Muerle, J. L. (1960). The Maxplane program for factor rotation to oblique simple structure. *Educ. Psychol. Meas.*, XX, 569–590.

[2] Dicken, C. F. (1961). Connotative meaning as a determinant of stimulus generalization. *Psychol. Mon.*, LXXV, 1–27.

[3] Harman, H. H. (1960). *Modern Factor Analysis.* Chicago: University of Chicago Press.

[4] Kaiser, H. F. (1959). Computer program for Varimax rotation in factor analysis. *Educ. Psychol. Meas.*, XIX, 413–420.

[5] Lawley, D. N. (1958). Estimation in factor analysis under various initial assumptions. *Brit. J. stat. Psychol.*, XI, 1–12.
[6] Osgood, C. E. (1962). Studies on the generality of affective meaning systems. *Amer. Psychol.*, XVII, 10–28.
[7] Osgood, C. E., Suci, G. J., and Tannenbaum, P. H. (1957). *The Measurement of Meaning*. Urbana: University of Illinois Press.
[8] Saunders, D. R. (1961). The rationale for an 'Oblimax' method of transformation in factor analysis. *Psychometrika*, XXVI, 317–324.
[9] Waite, G. (1961). A Study of Attitudes Using the Semantic Differential. M.A. Thesis. University of Birmingham Library.

Vol. XVI
Part I
The British Journal of Statistical Psychology
May
1963

NOTES AND CORRESPONDENCE

THE PSYCHOLOGY OF VALUE

To the Editor, British Journal of Statistical Psychology

SIR,—Those psychologists who hold that the essential subject of psychology is behaviour (and they now surely include the vast majority) will be grateful to Dr. Gregson for raising the general problem of ' value '. He thus takes us from the side-issues of art and aesthetics, on which too much of the recent discussion has been focused, to the central question of behaviour itself. Are some types of behaviour ' righteous ' and ' virtuous ' by their very nature, while others are ' vicious ' and ' sinful '? Or should we rather say that all are just adaptive reactions to the situations that confront us, but that some are more appropriate adaptations than others—more appropriate in the sense that they make for the survival and the progress of the individual concerned or of the social group to which he belongs.

Professor Burt and his co-workers want us to adopt the former view, alike in the spheres of aesthetics and of morals. " I am tempted to contend ", he writes, " that aesthetic relations, like logical relations, have an independent or ' objective ' existence: that the Venus of Milo would remain more lovely than Queen Victoria's statue in the Mall, the Taj Mahal than the Albert Memorial, even though every man and woman in the world had been killed off by an atomic war . . . A sonnet of Wordsworth, a finale from Wagner, a pile of apples painted by Cézanne—these convey more to us in five minutes than a scientist or philosopher could expound in an hour's lecture . . . The artist's search for beauty is but the special instance of a gift which lies within the reach of all—a gift which, far better than reasoned argument, will enable the young and uneducated (and in the face of the universe we are all of us uneducated and immature) to grasp the fundamental values."[1] All very alluring as a burst of rhetoric, but most unconvincing to the empirical behaviourist.

And just as Professor Burt and those who think with him believe in absolute principles of beauty, so they apparently regard the outdated ethical code of the Victorian moralists as absolute and universal. *The Young Delinquent* was no doubt a landmark in its day. But it is now sadly behind the times. Those who still follow its pre-behaviouristic principles are seeking to impose on delinquents and non-delinquents alike an immutable scheme of ethical commandments or norms which date back to Sinai. No doubt the parallel tables of percentages, scattered throughout the volume, and the carefully argued inferences based upon them, are highly informative. But as soon as Sir Cyril exchanges his professorial gown for a bishop's gaiters, he finds it difficult to make progress. He sermonizes at some length over the ' treatment and training ' of the potential criminal; and then, in a mellower mood, entertains us with a few racy case-histories that might almost have been picked from Dickens. It is only when he finally gets back to his quantitative studies of large samples, to tables of correlations and percentages, and to the illuminating bivariate frequency diagrams—the proper end and object of such researches—which show the relations between causal factors and behavioural traits, that he becomes his old scientific self.

Biologically and sociologically, however, honesty and the other virtues simply are "the best policy ', and nothing more. In war-time and for members of underground movements, dishonesty and deceit, theft, sabotage, and murder may often become virtues, in spite of the Twelve Commandments and the categorical imperatives of the neo-Kantian school. So far as moral judgements are concerned, surely Freud and Watson come far nearer to the actual facts and present-day requirements than Moses or Paul of Tarsus, or even Hegel and Kant. As Dr. Paterson has rightly pointed out, Professor Burt's assumption of an intuitive capacity whereby aesthetic and ethical 'values' (so-called) are directly apprehended in some miraculous fashion is an unnecessary and unverifiable postulate which modern psychology has rightly decided to drop. The various invocations of 'value' to lend a glamour to our traditional customs and conventions are simply a hang-over from a metaphysical and prescientific psychology, harking back to the days when psychology was a mere offshoot of theology—vestigial survivals as antiquated and as useless as the caudal muscles that once wagged the human tail. The man in the street will no doubt persist in expressing himself in terms of hallowed notions like 'righteousness', 'beauty', and 'truth', with 'happiness' and 'bliss' thrown in; but the psychologist, as a devotee of natural science, must scrupulously eschew all such transcendental concepts as likely to mislead him into a tangle of unrealistic inferences.

---

[1] *How the Mind Works* (2nd ed., 1948), ch. XV, ' The Psychology of Art '. Much the same view is defended in regard to ' ethical values ' in ch. XVI, ' The Psychology of Religion '.

After all, what are the aims of natural science? Primarily, as Watson has reminded us, "to predict and to control". With this twofold end in view it starts by observing and recording, and by measuring and calculating; and it confines its conclusions to "explanation in physico-chemical terms" (*Behaviour*, p. 26). This, as Dr. McLeish has admirably argued in his recent book, is a workable plan of campaign which has served us well; once we depart from it we get bogged down in a metaphysical morass.

Let us imagine Professor Burt standing before the Venus of Milo in the Louvre. The statue, he tells us, is in his view a thing of beauty—far more beautiful than the Edwardian pile before the Palace; and, sweeping aside all modern inquiries, he resurrects the old mystical doctrine about the permanent and independent existence of beauty and all other forms of value. One would have thought the sheer antiquity of the doctrine was sufficient to condemn it. However, let us take the contention at its face value, and ask what precisely it amounts to. Simply that Professor Burt finds pleasure in gazing at the Melian Venus, and none whatever in gazing at Queen Victoria in the Mall. And then, by a figure of speech he projects his pleasure onto the marble object, and describes the statue itself as pleasing, or, in his terminology, 'beautiful'.

And how does our up-to-date behaviourist describe the occurrence? Looking at the statue and experiencing pleasure, he reminds us, are processes directly dependent on certain physico-chemical changes in the spectator's sense-organs and nerves; and a state of pleasure is a really transient physiological condition in which the smooth and unimpeded flow of the nerve-currents in the brain are gradually restoring the organism's previous phase of equilibrium, which the visual stimulus has disturbed. For the scientist, therefore, pleasure is wholly reducible to certain physico-chemical processes; and hence cannot be an attribute of the statue, much less hypostatized as a 'basic element in absolute Reality.'

According to the behaviourist, the task of the psychologist is (to quote Watson once again) "to describe, not to prescribe", for psychology is "a positive and not a normative science". Its working procedure is to postulate plausible hypotheses based on a minimum of assumptions; and, in order to establish those hypotheses, what the psychologist has to do is to construct operationally testable predictions of human behaviour, i.e. forecasts based on regression equations or their equivalent which shall achieve a maximum amount of accuracy. In his work on factor analysis this is all that Burt professes to attempt; but in his theories on aesthetics, and above all in his pronouncements as an educational, industrial and criminal psychologist, he goes on to lay down normative rules. To justify those rules he invents an invisible realm of ethical and aesthetic 'values' existing in and for themselves, like Plato's 'world of pure ideas'; and he thus continually drags in concepts and postulates long ago repudiated by objective psychology. Is it not time therefore to insist that such questions can only be scientifically investigated, by treating them as we treat all the other problems that arise from the conduct or predicaments of our fellow human beings—namely (as Dr. Gregson has rightly contended) *behaviourally and statistically*.

I suggest we should begin by considering how the socalled ' sense of value ' seems to develop in the tiny child. Tommy, when only 4 months old, will cling to his doll, and resist all attempts to withdraw it. At 6 months he will clutch at any attractive object that catches his attention—a coloured ring, a dangling rattle. At 12 months he will point at, and later he will call for by name, whatever appeals to his passing fancy. Soon he develops a pride of possession and a sense of ownership. As he learns to run about and talk, new desires get voiced. " I want to win that ally-taw," he shouts. Now the doll, the ring, the toys that he owns, the ally-taw he wants to win—each in their own way have thus acquired, temporarily at least, what the high-brows would call ' intrinsic value '. Gradually he picks up a vocabulary to describe these values: at 18 months things are " nicey-nicey ", later they are " lovely ", or " grand ", or " ripping "—the first vocal steps towards aesthetic appraisal. And then? Unfortunately for later ages there is a lamentable dearth of developmental research.

A genetic study of this sort is surely far more informative than a theoretical discussion of the alleged use of the Fechnerian golden section in art which in any case the experimental studies of twentieth century aesthetic psychologists have completely rebutted. Professor Burt appears to find confirmation of Fechner's theory in the products of Roman and Renaissance architects. Whether that be so or not, no present day artist or architect would take seriously these exploded speculations of the socalled ' founder of experimental aesthetics '.

Piaget's researches show that " moral judgements and the sense of moral values " develop in much the same way as I have outlined for aesthetic judgements: only in the case of morality the influence of the social milieu—of parents, playmates, teachers, and schoolfellows—is more conspicuous. A " nice " or " lovely " marble is the kind of marble Tommy at the time of speaking himself prefers; " nice " or " good " behaviour is the kind of conduct that pleases Tommy's mother, and, if he is a " good " boy, Tommy likewise obediently prefers. A phrase ostensibly attributing aesthetic value, e.g., " That's a nice shade of red ", as Dr. Gregson points out, is nothing more than " a terse way of referring to a consensus of choices ". By a like train of reasoning we can equally conclude that ' ethical ' judgements, such as " That's a nice way to behave " or " That's the proper conduct for all good citizens to adopt ", are also nothing more than " a terse way of referring to a consensus of choices " between alternative types of conduct.

Professor Burt would have us treat aesthetic and ethical values as independent attributes of real events and objects, indeed of reality itself. But this surely is just the naïve and uncritical impression of the romantic moralist. That beauty and goodness actually exist has of course been maintained by numerous divines from Augustine to Dean Inge; but in their case the independent existence of value is merely an archaic theological dogma deduced from, and dependent on, the traditional belief in God. At best, value in any case could only be value in or for a person, and therefore essentially subjective.

Modern behaviourism, profiting by the biological standpoint introduced by Darwin and Huxley and elaborated by Watson and Freud, has shown that all these so-called ' ethical ' and ' aesthetic ' principles are merely the adaptive behaviour of a gregarious and highly evolved primate—behaviour which in the course of ages has proved requisite for the survival and progress of his race, and which is learnt and relearnt by the conditioning of each individual afresh within the social group into which he is born. The moral progress and the high moral standards on which we plume ourselves today are based, not on the ' eternal verities ', but on the results of a struggle for survival between different tribes, nations, social classes, and religious sects: they are biological utilities resulting from biological evolution, not ethical values resulting from increased moral insight. The chief problem, as Freud has shown, is to reconcile the impulses of the *Id* with those of the *Superego*—the vital with the social: thus the 'good' or 'virtuous' life is simply a life in which all our impulses are harmonized.

If we adopt this simple interpretation, then we can follow Bertrand Russell's advice, and apply Occam's well known maxim—the principle of minimum assumptions. We can throw overboard once and for all this needless cargo of 'absolute goodness', 'absolute beauty', and all the other 'values'—the stock in trade of the parson in the pupilt and of the old fashioned literary critic; and so eventually come to rest on the dry *terra firma* of the physical and biological sciences.

What after all, in plain language, do people mean when they talk about 'value'? Value is just a label or price-ticket affixed to an object to indicate that it is, temporarily at least, an *object of desire;* and desire in turn is itself just another name for certain motor responses, overt or incipient, which lead, when finally consummated, to the acquisition of the object, as in the 'desires' we call hunger or sex. There is nothing occult or mysterious about such tendencies. Nor is there the smallest need to invoke obsolete teleological concepts such as 'purpose' or 'will'. The responses are primarily no more than reflex tendencies; but in human beings they are subject to widespread conditioning, in a manner which conforms to certain principles or patterns that have been experimentally demonstrated. When overt (as in lower animals), the basic tendencies can be readily observed; when merely incipient (as so often in human beings), we may, as Dr. Gregson says, have to rely on our subjects' "verbal reports". In either case all the relevant data can thus be satisfactorily described in terms of the accepted concepts of orthodox behaviourism; and the retention of specific ethical or aesthetic notions becomes superfluous, and, should, as Dr. McLeish has so convincingly argued, be forthwith jettisoned as involving a multiplication of entities beyond what is absolutely necessary.

There is in fact but a single basic 'desire'—a single goal or end-state towards which all our motor responses tend. Biologists, behaviourists, psychoanalysts, and cyberneticists—Cannon, Freud, Hull, Ross Ashby—each working independently along his own special line—unite in teaching us that what determines an organism's behaviour may be reduced to one simple homeostatic law—*the recovery of stable equilibrium.* The change in the environment which constitutes the 'stimulus' disturbs the organism's pre-existing state of equilibrium, and the organisms 'response' consists in a series of adaptive changes, whereby (so far as possible) the original condition of equilibrium is (with minor adjustments) once again reinstated. Pain or rather unpleasure is just a name for the disturbed stage (the 'goal' to be reached, the 'value' to be realized). Survival itself is only another name for preserving or restoring the original condition of stable equilibrium we call 'life'. That is all that the behaviourist means by his basic principle of 'need reduction', and that all that the psychiatrist implies when he tells us that 'the fundamental requirement is security'—yet another synonym for stability.

For both the research-worker and the theorist the outcome of the whole discussion may, I think, be summed up in three pieces of advice. First of all, problems of 'value' (socalled) have turned out to be essentially psychological problems, not philosophical; and yet they are still sadly neglected by the empirical psychologist. Secondly, since they are psychological problems, they must be interpreted in terms of the concepts of present day natural science, not of an obsolete metaphysics, and must be investigated by the methods of natural science, that is to say, by experiment and statistical analysis; thirdly, before any such investigations are planned or undertaken, the first stage must be to review the few rival theories that are left, and then formulate a plausible down-to-earth hypothesis, which can be tested operationally by an objective study of actual behaviour.

J. E. Whittaker.

SIR,—In her reply to Mr. Paterson Miss Williams confines her evidence almost exclusively to factorial data obtained from a statistical analysis; and yet the conclusions she draws plainly imply a background of psychological or philosophical assumptions which she nowhere explicitly formulates. The individual whose order correlates most highly with the order indicated by the 'general factor' is, so we are told, the individual with the highest capacity for aesthetic appreciation. But is it the business of the psychologist to lay down a fixed and absolute standard of taste—a rather dubious standard, it must be confessed—as a canon to which everyone should conform? Professor Burt adopts the same procedure in his tests for artistic ability, and a very similar method in assessing the moral character of young delinquents. And with this is conjoined a fanciful notion of art and music as curative exercises (*The Young Delinquent*, p. 523). To me the whole notion sounds like an attempt to revive the psychological mysticism of the gnostics and the Jewish cabbalists —the ancient doctrine (as a leading Rabbi puts it) that " beauty is a part of, and indeed the key to, the nature of the universe—a door opening on the lost spiritualism which modern materialism, like the materialism of earlier days, has forgotten and left behind ". Will Miss Williams or Professor Burt tell us whether this is what they want to resuscitate, and if not where it all comes from?

Most present day psychologists would probably endorse much of what Dr. Gregson and his fellow critics have said about ' beauty ', ' goodness ', and other forms of ' value ' (this *Journal* XIII, pp. 87–88, XIV, pp. 72–75, and XV, pp. 163–5). Nevertheless I feel that they have not really proved their case, but merely stated it. As things are, though the problem is essentially psychological, it is—unexpectedly enough—not to contemporary psychologists, but to contemporary philosophers that we have to look for the clearest and most cogent analyses of the relevant psychological phenomena and the various theoretical interpretations of the facts—to Bertrand Russell, Gilbert Ryle, and above all A. J. Ayer. Indeed in Professor Ayer's ' emotive theory of value ' we have what in my humble opinion is really the final word on the subject. It seems even more convincing than Dr. Gregson's simple hedonistic theory. Nor have Professor Burt and those who think with him attempted to reply. I need not recapitulate the various arguments here; they are set out with admirable clarity in Ayer's *Language, Truth and Logic* (pp. 109f.).

But I write chiefly to emphasize two further points. There are a pair of obvious objections fatal to any theory that would make value self-subsistent and objective characteristic existing independently of all human recognition: first, the frequent disagreement between different persons in regard to specific instances; and secondly, the fact that the aesthetic quality of an object (or the ethical quality of an action) is itself dependent on other qualities that are acknowledged to be merely secondary and subjective.

In all other fields we adopt a simple criterion. When a perceptible property varies only with variations in the different objects and not with variations in the different perceivers or in the different conditions under which it is perceived (as for example, the height or width of a physical object, or its geometrical form or shape), then we regard it as objective and real; on the other hand, when for one and the same object, the property varies with the perceivers or the conditions, then we regard it as subjective. In music, poetry, and painting it is notorious that what one man hails as beautiful, another spurns as detestable. This is most obvious in judgement on *modern* art and *modern* music—a rich field of study which psychologists have so consistently neglected. Similarly, in morals: the church still considers perfect chastity the ideal; *Sixth Form Opinion* recommends premarital intercourse as a wise preliminary to formal marriage.

The second objection is still more conclusive. The appreciation of a picture depends essentially on the pattern and the harmony of the colours; the appreciation of a piece of music on the pattern and harmonies of the sounds. But colours and sounds are wholly subjective —mental contents rather than the attributes of objects. Once again therefore it follows that the beauty which depends on these must also be subjective. " Beauty ", as Hume observed, " is in the eye of the beholder."

Indirectly both these arguments could be used to confirm the hedonic theory of beauty which Dr. Gregson, if I understand him rightly, prefers to advocate. It is the theory which the whole psychology of aesthetics from Fechner to Valentine seems to support. Beauty is just a popular name for the pleasurable hedonic tone which suffuses certain intellectual experiences and our emotive reactions; and those individuals who rank high in the 'general factor aesthetic capacity' (as Miss Williams terms it) really do so because by their preferences they indicate that they are the most sensitive to this type of pleasure.

In the effort to extricate himself from the impasse to which these arguments inevitably lead, Professor Burt introduces a newfangled distinction between 'explanation' and 'proof'. If, it appears, the door of 'proof' is shut, then a loophole of escape can be provided by following the line of 'explanation'. But what precisely is the difference between the two? Surely, if an explanation cannot be proved, it is worthless even as an explanation.

At times Professor Burt himself seems to come down on the subjective rather than objective side of Dr. Gregson's metaphysical fence. Each of us, he says, man, woman or child, "likes that kind of picture, poem, or melody which is sufficiently intricate and diversified to keep his attention continuously active, and yet shows a sufficient uniformity of plan not to defeat entirely his efforts at attentive comprehension," (*How the Mind Works*, p. 304). Consequently, as we are repeatedly told by those who have made a systematic study of aesthetics, a composition, whether musical or pictorial, must have a latent scheme of organization; it must display symmetry, balance, and proportion. There must be variety, but the variety must be integrated into some kind of unity. Hence too Fechner's emphasis on such quasi-mathematical properties as simple ratios, golden sections, equilibrium, and the like. Now, so far as the material object is concerned, these are just metrical characteristics. They produce the effect of beauty because, according to Burt's theory, they at once stimulate and facilitate attention. Thus, if beauty does not reside in the eye of the beholder, it surely resides in his capacity for attention. And that is no less subjective.

SAMUEL J. LEVI

SIR,—Neither Dr. Gregson nor Dr. Levi appears satisfied with the brief answer which I gave to Mr. Paterson's inquiries. Dr. Gregson (see this *Journal*, XIII, p. 89f. and XIV, p. 73) supposes that I have been "misled . . . because the problem is not so simple as the ordinary nature-nurture dispute". I agree about the complexity of the problem. I ended by referring to the evidence for innate capacities merely because that was an issue raised by Mr. Paterson. I also agree with Dr. Gregson over the importance of "differentiation between hedonic responses and responses recognizing formal properties as present", for what he describes as 'formal properties' I should consider to be conditions necessary for 'aesthetic value' or 'beauty', as Professor Burt has already argued. But the comments Dr. Gregson then makes on the distinction are decidedly puzzling. "The absolute theory", he writes, "would permit us to say that the subject whose responses identifying works of art are most typical will be the subject whose personal preferences are most closely in agreement with what he identifies as good." Now the correlations I reported are in complete accord with this corollary. So far therefore the 'operational test'[1] confirms the 'absolute theory'. Yet in the end (*ibid.*, XV, p. 165) Dr. Gregson appears to reject "the notion that value is absolute" altogether.

---

[1] The phrase is Dr. Gregson's (*loc. cit.*, p. 163). It has long been a highly popular cliché. But I could wish he had enlightened us as to the significance of the adjective. Supposing that a 'prediction' was 'testable' though not 'operationally' testable, and that it actually passed the test, would Dr. Gregson then reject it? In the following paragraph there is another statement I fail to understand. He tells us that he intends to "indicate where he thinks we *might* treat of value as though it were objective, and where we should *not* do so". Yet his final conclusion is that "all the above seems to me acceptable *without any* idea of value as objective" (my italics).

Dr. Levi complains that I confined my evidence almost exclusively to " factorial data " obtained from a statistical analysis, and yet the conclusions drawn " plainly imply a background of psychological or philosophical assumptions which are never explicitly formulated". That I readily admit. But I was replying to a specific request for ' actual factorial data ' (*loc. cit.*, p. 88) and most readers apparently suppose (not quite correctly perhaps) that the statistical journal is the place for statistical evidence and the general journal for the discussion of more general assumptions. In point of fact the theoretical background had been set forth elsewhere by Professor Burt, under whose guidance most of these collaborative studies were carried out. Nevertheless I for one would heartily agree that it may be extremely misleading to be given statistical data *without* at the same time formulating and defending the underlying assumptions. Indeed, as the editor reminds me, that is precisely why this particular section of the statistical journal has been introduced.

This being so, may I now reply more fully to the criticisms which Dr. Gregson and other correspondents have advanced? As a result of the experiments described in my reply a standard order was (as Dr. Levi says) compiled by taking the average of the various rankings, weighted according to factorial principles; and the examinee's capacity for artistic appreciation was assessed by correlating his ranking with this standard order. Now both the purpose and the assumptions involved in this procedure have been widely misunderstood by non-psychological art-critics. It is objected that the whole notion is vitiated by invoking a " dubious standard of good taste " against which the results can be assessed. Dr. Levi denies that there is any such standard. Similarly Dr. Gregson apparently supposes the ' statistical psychologists ' who are responsible for the method thereby assume " that [artistic] value is absolute and not relative "; whereas owing to " the ephemeral nature of the test material " the norms adopted must be " culture-loaded ". I should not have thought that the paintings of Raphael, Leonardo, etc., were " ephemeral "; and, since they are highly rated by persons from widely different cultures— Eastern as well as Western —there is little evidence that the ' culture loading ' is so heavy as to invalidate the procedure. But these are side issues. The real point is that the argument adopted by Dr. Gregson and the other critics fails to distinguish between beauty and assessments of beauty. Their contention is thus tantamount to arguing that, because the *assessments* are not ' absolute ', but ' relative ' to the view of the group of judges and in that sense ' subjective ', therefore *beauty* itself cannot be ' objective ', but must be ' subjective '. We recognize that the standard order which forms the ' norm ' may be to some extent relative and even culture loaded, and that the ' standard of taste ' may in some measure change from place to place and from age to age; but all the evidence shows that the changes are far smaller than is supposed by our critics. Moreover, in actual practice the test undoubtedly serves its essential purpose (though no doubt, like all tests, somewhat imperfectly). What we want to deprecate, however, is the notion that, in arriving at this provisional order, the ' statistical psychologist ' is thereby *laying down an absolute norm or standard*, and so " prescribing what the standard of taste should be ". The so-called ' norm ' (it is not our term) is put forward merely as a factual result, not as a " canon to which every one should conform ". Dr. Levi revives the old argument about differences in taste. " What one man hails as beautiful another dismisses as detestable." But what is the evidence for these ' notorious ' differences? If the phrase " one man " refers, as the context suggests (e.g. his allusions to ' Sixth Form Opinion '), not to the specialist or the expert, but to the ordinary man in the street, why should we worry about what such a person thinks of this or that painter poet, or music, any more than we worry over his views about flying saucers, or seek his opinion about the assumptions of Euclid or Einstein regarding the nature of space? Surely it is the verdict of the expert and the specialist that counts.

I fully sympathize with the fear of " psychological mysticism ", which is so often associated with the doctrine that " beauty is a part of and indeed a key to, the nature of the universe—a door opening on the lost spiritual world which modern materialism is supposed to have forgotten and left behind ". But it can hardly be maintained that the " statistical psychologists" whom Dr. Gregson and Dr. Levi have singled out for criticism intended their results to support the extremely naïve and far-fetched speculations that are so popular in this country and America.[1]

However, Dr. Gregson's present position is that "we have now moved to the wider issue of the status of value in psychology", i.e., of value in all its forms. And much of Mr. Whittaker's recent letter is consequently taken up with a flat denial of value in this broader sense—in particular of ethical value, and I presume of logical value, as well as of aesthetic value. He makes no attempt to counter, with any first hand facts or evidence, the almost universal impression that values are indeed objective realities that we directly perceive; instead he reduces value to what is desired, and what is desired to what gives us pleasure; what gives us pleasure, we are then assured, is that which in the course of evolution makes for our survival and progress, and consequently meets with an emotional approval from society. All this, I confess, sounds not so much like an alternative theory, as a hotchpotch of theories.

Moreover, although he professes to refute the notion of value in all its main forms, his actual arguments, like those of most other psychologists who have registered a similar denial, refer chiefly to aesthetic value—i.e., to beauty. Apart from a small handful of pragmatists (who have no representatives, so far as I know, among contemporary psychologists) few have attempted to deny the objectivity of logical values, i.e., the validity of rational arguments, or in one word, Truth. And apart from the bolder Freudians and the earlier and more defiant behaviourists, not many psychologists have ventured *openly and explicitly* to deny the objectivity of ethical values—though several eminent philosophers have dared to do so. Our present critics seem to keep chiefly to beauty, because, of the three traditional types of intrinsic value, beauty offers the easiest and least controversial case in which their contention can be maintained. And yet, to the practical psychologist—those, for example, who have to deal with delinquents, neurotics, and the numerous individuals who suffer from moral or religious conflicts—it is the question of ethical values that is most urgent.

---

[1] Dr. Levi very naturally inquires where such speculations come from. They are due, I suspect, to the journalistic confusion between the views of psychologists and the views of certain psychiatric writers ("medical psychologists" as they are popularly termed). In recent times the chief psychological support for the various notions to which he refers seems to have come from the followers of Jung: cf., e.g. H. Westman, *The Springs of Creativity* (Introduction by Sir Herbert Read), and H. Reich, *Seelenbilder: Strukturanalyse der Seele durch gegenstandsfreie Bilder*. Strange to say, even eminent art critics, like Sir Herbert Read, and equally eminent men of letters like Mr. Aldous Huxley, appear at times to share these more extravagant notions. "Creative art", we are told, "and the vision of beauty are the surest cures for the mental ills of today . . . Art in its widest sense has become almost the sole outlet for those constructive energies that enable the psyche to condition itself to an objective harmony" (H. Read, *A Letter to a Young Painter*). There is of course undeniable evidence, published and unpublished, showing how painting, or attempts at painting, may, occasionally at least, assist in the cure and rehabilitation of psychoneurotics, and even at times throw light on the nature of their disorders. Unfortunately, those psychiatrists and art-teachers who wish to demonstrate the merits of such methods never attempt to offer statistical evidence for their views (such as would be required to demonstrate the value of a new drug); they take refuge in semi-mystical arguments about 'archetypes', 'health-giving Mandalas', 'Ursymbolen', and 'Urgrundformen', which fully deserve Dr. Levi's scornful comments. At the same time an esoteric air is commonly imparted by wrapping up plain and plausible statements in the gnomic jargon of so called 'depth psychology': (for further instances see Anton Ehrenzweig's *Psychoanalysis of Artistic Vision and Hearing* and Adrian Stakes' *Painting and the Inner World*; but even in the latter such a sentence as "Nudity, an absolute so often achieved in an instant without rebuff, furnishes no reference to the endless mind" defeats any attempt to give it a clear cut meaning). The concepts and the lingo are certainly derived very largely from 'theosophy' —not so much the 'theosophy' of Boehme, Schelling, and Swedenborg (though indirectly this has supplied a good deal), but the dubious brands associated with the names of Madame Blavatsky in the U.S.A. and Mrs. Annie Besant over here, and (perhaps most influential of all) Rudolf Steiner, though he preferred to christen his version 'anthroposophy': see for example L. D. Ettlinger, *Kandinsky's 'At Rest'* (Charlton Lectures on Art), Rudolf Schott, *Der Maler Bo Yin Ra* (whose original name was Josef Anton Schneiderfranken), and Gladys Mayer's *Mystery-Wisdom of Colour*. I myself would freely admit—what the behaviourist would prefer to deny—that even the ordinary phenomena of consciousness involve the deepest mystery; but we need not insist on treating mystery qua mystery as something sacrosanct *beyond* our everyday life instead actually part of it.

Now here the theoretical psychologist has proved most unhelpful. To begin with, he tends to shun the whole topic of moral value. There have been plenty of researches on the psychology of aesthetics—and even on the psychology of thought and reasoning; but hardly anything—at any rate during the last thirty years—on the psychology of ethics, of moral judgement, and of moral standards. Secondly, after the first world war the writers who had the most powerful influence in determining the trend of psychological thought were the writers who shared the popular passion for debunking. Indeed in the public mind, which to a large extent is merely a reflection of the journalistic mind, Freud with his doctrines of sex, the super-ego, and religion as an 'illusion', is held to have successfully debunked the traditional moral codes; Watson and the behaviourists are similarly held to have debunked the traditional notions of consciousness, of a mind or soul, and of everything approaching purpose or free will; the sociologists with their theory of 'culture-loaded' criteria (the phrase borrowed by Dr. Gregson) are held to have utterly exploded the notion that values of any sort are objective or real.[1] Thus indirectly, I believe, Watson and Freud and their modern followers, particularly among the clinical psychologists and psychiatrists who deal with youthful offenders or seek to guide public opinion, are themselves in part responsible for the marked decline in moral standards among the rising generation: too often they seem more ready to condone what is evil than to condemn it. Thirdly, as has been frequently remarked, in actual practice both psychologists and psychiatrists, when dealing with juvenile delinquents, "lean over backwards" (to adopt fashionable mteaphor) in their efforts to disown any attempt to impose or bolster up traditional standards, or to advocate moral ideals: they seem scared lest this should make them sound unscientific. This non-committal attitude is clearly apparent in the reports of contemporary researches on delinquency (see, for instance, *The British Journal of Criminology*, *passim*). Hardly any investigator appears to ask the offenders he is studying the two most obvious questions: 'Why did you commit that offence?' and 'Do you, or do you not, consider what you did to be wrong?' Few of them are willing to recognize that most petty thefts are deliberately committed for gain, or that most petty thieves are ready with plausible justifications and not unnaturally conclude that the psychiatric interviewer himself privately believes the whole notion of absolute moral standards is sheer humbug. The vogue at the moment is to attribute such actions to quasi-pathological and half-unconscious motives, to ambivalent father-fixations, to the morbid effects of a broken home, or to the possibility that the delinquent was 'emotionally disturbed' or suffering from some 'maladjustment' for which he was not to blame or even from a 'true criminosis' (whatever that may mean)—anything rather than recognize that he is morally responsible. Nearly every psychologist seems agreed that the State should not administer punishment by way of *retribution* for the crime, i.e., to vindicate moreal standards, first because "the members of the State are rarely unanimous about the wrongfulness of particular actions", and secondly because both psychoanalysis and behaviourism postulate a "strict determinism", which is wholly incompatible with the conclusion that the wrongful act might result from free and voluntary choice. Deterrent punishment and reformative punishment they are ready to allow in theory, since these are consistent with the principles of 'conditioning' and 'reinforcement'. Yet in practice the main tendency is to deprecate all form of punishment in favour of quasi-psychiatric 'treatment' (which in point of fact is seldom attempted even in the rare cases where it is manifestly needed).[1]

---

[1] I have been asked to give instances of actual pronouncements made by psychiatrists, psychologists, or sociologists which bear out these allegations. Three must suffice. In setting out his "plea for a purely rational basis for cultural laws, that is to say, for deriving them from social necessity", Freud assures us that "along with their pretensions to sanctity, the rigid and immutable nature of our moral laws and regulations would also cease: men will realize that these have been made not to govern them, but to serve their interests" (*The Future of an Illusion*, p. 73). Flugel has maintained that the general trend of modern psychology has been towards "the substitution of a psychological for a moral point of view"; this, he admits, has aroused "a considerable fear that psychologists in their attempts to understand the motivations underlying values—ethical, aesthetic, or religious—are thereby in danger of destroying these values, or are even actively seeking to destroy them"; but he defends the whole movement on the ground that "it promotes better adaptation to the social environment" (*Man, Morals, and Society*, pp. 10, 259). Finally, let the reader glance at the summary in Table 2 in Kimball Young's *Handbook of Sociology* (more particularly item 6, "growth of self, personal goals, rights and duties") and note the implications of the conclusions in the columns headed 'Germans' and 'Americans' respectively.

The most recent as well as the most outspoken statement of the behaviourist position is that contained in Dr. John McLeish's recent book, *The Science of Behaviour* (1963). It has already earned high praise from at least one eminent professor working in the field of personality. McLeish's conception of psychology, like that of Watson himself, is based essentially on the Pavlovian approach. "Pavlov's theory", we are told, "is a model theory for the science of behaviour." In place of the older dualism, which recognized both conscious processes and material processes, Pavlov, we are told, has taught us to substitute a materialistic 'monism'; what is subjective must be ruthlessly rejected in favour of what is purely objective. And in this way psychology is at last "freed from the Procrustean framework of introspection and theology". The cardinal error of all "theological explanations of behaviour", it is argued, was to "posit a freedom of choice on the part of the behaving subject." By making "the principle of determinism basic" behaviourism has succeeded in relieving us of all this "arbitrary hocus pocus".

Chapter II of Dr. McLeish's book works out the corollaries of this doctrine for the problems of 'sin, crime and criminal responsibility'. Since "the whole theory and practice of our criminal law is built up on the assumption that the criminal *intends* his action," and "intention depends on free choice", our current attitude, both popular and legal, must be drastically changed. Dr. McLeish admits that "more recently the quantum physicists have suggested that science must discard the notion of cause"; but (so he assures us) "this suggestion is not regarded favourably by scientists"—an amazing statement in view of the fact that all except one or two unorthodox critics now accept "the principle of indeterminacy". He recognizes too that there are still a number of old-fashioned moralists who "assert that this [behaviourist] theory is destructive of law and morals;" and as a result "psychologists are now often accused of weakening the moral fibre of present day society". Nevertheless, in psychology as elsewhere the causal law is inexorable. The archaic "religious assumptions" on which the popular notions of mind, will, purpose, moral values, and the like have been based "are excluded": they "merely multiply entities which are unnecessary in the explanation of behaviour". And thus our final conclusion must be that "the whole idea of sin and delinquency is alien to psychology".

Whether there may be any actual facts or data to support or explain the popular notions of 'intention', 'consciousness', 'purpose', 'value', etc., Dr. McLeish does not stay to inquire. It is curious that a writer who insists so strongly on 'objective' methods should decide these particular issues simply on grounds that are flagrantly 'subjective'—namely, his own personal preference for two 'basic assumptions', first, 'the canon of simplicity', which enables him to dismiss in advance anything that does not fit his own simple-minded theory as 'wholly superfluous', and secondly the principle of 'causation', which is itself (as Russell and others have repeatedly pointed out) "one of the most subjective notions that have ever found their way into science". It may perhaps seem wrong to blame Dr. McLeish for these rather dogmatic pronouncements, since according to his hypothesis his own behaviour must also have been ' automatically determined '. Yet, strange to say, having spent so much time on insisting that the notion of ' intention ' must be eliminated as contrary to universal determinism, he ends on the very last page of his book by explaining what has been " the intention of the author ". If then authors—or at any rate this author—can have ' intentions ', why not criminals?

---

[1] See P. Callard, 'Punishment by the State', *Brit. J. Delinq.*, X, 1959, pp. 36f. For the term 'criminosis' see *ibid.*, X, 1960, p. 175.

Mr. Whittaker's approach (pp. 48f.) is avowedly the same as that of Dr. McLeish. On the question of criminal responsibility and kindred problems the views adopted in *The Young Delinquent*, and by " those who still follow its pre-behaviouristic principles, are " he tells us, " sadly behind the times ". However his attack on the conclusions set forth in that volume is evidently developed, like that of Dr. McLeish, purely from the abstract standpoint of the psychological theorist. May I therefore, on the basis of a fairly long experience both of school children and of work in juvenile courts, try to answer his criticisms from the practical standpoint? Dr. Whittaker argues that those who accept the principles assumed in the chapters he criticizes " are seeking to impose an immutable scheme of ethical commandments or norms dating back to Sinai ".[1] This seems to be the same type of misunderstanding that he shows in discussing so-called ' aesthetic norms'. In point of fact I can find no such doctrine in the book he has attacked; and in spite of all his protests it must surely be acknowledged that its two main conclusions have since been amply substantiated by actual events. The first and most important result was the demonstration of the ' plurality of causes ', (p. 590). This, strange to say, is still ignored by the vast majority of current researches, which endeavour to establish some single, highly vague and generalized factor as the essential cause. The choice of the factor varies from one investigation to another; but all have one thing in common: they scrupulously avoid any ethical reference, and tend to substitute a pseudo-medical or quasi-sociological terminology for plain moral concepts. The second conclusion was itself twofold, namely, that " *material* conditions, such as poverty are far less important than *moral* conditions " (p. 607). When the book was published, nearly every criminologist of the day laid the greatest stress on material conditions; by abolishing poverty, it was argued, the welfare state, when at last it arrived, would go far towards abolishing delinquency and crime; and Burt's conclusion was denounced by contemporary reviewers as " contrary to all experience ".[2] He on the other hand laid chief emphasis on psychological factors, and most of all on the lack of any guiding purpose in life.

---

[1] To avoid this type of criticism may I say that those of us who took part in the investigations fully recognized that one of the main difficulties confronting us was precisely the way in which the nominal code was out of step with the rapidly changing attitude of the times—particularly in regard to such matters as parental authority, sexual ethics, homosexuality, suicide, and the like? However, as Burt warned us in advance, " the psychologist's duty is not to echo Cicero and the Ciceronian laments of the pulpit with an ' O tempora! O mores! ' but to set about collecting evidence to see what the changes actually were."

[2] " Poverty," he argued, " much as we rightly strive to abolish it, often provides, not a motive for crime, but wholesome incentive. The youngster from a poor home knows he must work hard and conscientiously if he is to gain and keep a job; and this ever-present knowledge at least gives him a purpose in life. Indeed the need to labour, the curse pronounced on Adam has proved one of man's chief blessings." My own experience in following up pupils after their school days are finished has fully convinced me of the truth of these remarks. While I am writing I have in front of me a note from an employer who observes that, " as the risk of unemployment has declined, and as every youth has become assured of a steady and indeed a generous wage, and learns that laziness need not mean dismissal (since his fellow-workers will at once down tools till he is reinstated), his one and only object is, not to accomplish a good job and to uphold the credit of his firm, but to do a minimum of work for a maximum wage. As a result his day-to-day existence has gradually degenerated into a weary search for whatever may take the place of a higher aim—the excitement of betting or gambling, a round of petty sexual adventures, and for the rest sheer boredom. The reduction in poverty, at any rate in this town, has certainly not meant a parallel reduction in vice and crime ".

All this of course was in direct contradiction to the inferences drawn from behaviourist theories. If in keeping with behaviouristic theory we deny that moral values have any independent reality, or can be directly recognized as intrinsically valid, then life becomes essentially aimless. Instinct and emotion, the selfish desire for pleasure, for fun, for thrills, and for a good time generally, remain the only objects which the undisciplined child or teenager can be expected to seek. If Mr. Whittaker wants to have this proved by statistics, he will find plenty of confirmation in the book.

I fully agree with Mr. Whittaker that the primary need is 'psychological research on an objective statistical basis'; but, unlike Mr. Whittaker, I think that case-histories must furnish the requisite data' I agree too that the first problem to which such research should address itself is "the nature of so-called values"; and for that reason I feel it odd that no one has taken up a suggestion, thrown out in the very chapter he criticizes, that an illuminating study would be a comparative investigation of a group of exceptionally *virtuous* members of society, side-by-side with a control-group: "the real puzzle", so the author tells us, "is not so much why some people are wicked, but why so many are good". As for 'treatment and training', systematic research is here still more acutely needed; and the most profitable type of treatment for the statistical psychologist to investigate would, I venture to suggest, be not the biological, or the sociological, or the psychiatric methods that are so widely advocated today, but the merits of a specifically educational approach—a study of moral training, which should begin not after the offender has been caught and convicted, but from the very earliest years.

However, as he himself implies, before any investigation is started, we need more preliminary analyses, detailed, systematic, and comprehensive, both of the various concepts involved and the alternative theories available. There have been quite recently one or two experimental inquiries dealing with 'values'. To evade behaviouristic criticism they are usually described as studies, not of judgements of value, but of 'evaluative behaviour'. And, since behaviourism has swept away all interest in traditional theories, the investigators have been left to improvise their own rough and ready procedure. The usual method is just to throw together almost at haphazard a number of 'verbal labels'; the subjects' 'choice-behaviour' is then arbitrarily assessed; correlations are worked out, and the coefficients are factorized to see whether or not one of the rotated factors may prove to be an 'evaluative factor'. Needless to say, no introspective checks are applied to confirm the proposed interpretations of the factors. Certainly such an experiment is 'objective psychology' of a sorts; but it can hardly be regarded as scientific research. Let us hope that British investigations, with a more fruitful tradition of statistical research behind them, will plan something a little better than this.

E. D. WILLIAMS

Vol. VXI
Part 1
The British Journal of Statistical Psychology
May
1963

THE PSYCHOLOGY OF VALUE

A Reply

By Cyril Burt

> Sternel la voce del verace Autore
> ' Io te farò vedere ogni valore '.
> Dante, *Paradiso*, XXVI, 40–2

I. Are Values Objective?

The Multiplicity of Views. Oscar Wilde once announced as the title of his lecture: "Are the Critics of *Hamlet* Really Mad, or Only Pretending?" In a similar spirit one is sometimes tempted to ask, "Are the psychologists who criticize the notion of moral value really as unprincipled as they sound, or are they only pretending?" From the comments of several first-hand observers who have written to this *Journal*—educationists, social workers, magistrates in juvenile courts, as well as from the columns of the daily papers, it is plain that there is a widespread belief that (in the words of one of them, p. 55 above) " the modern followers of Freud and Watson, particularly among the clinical psychologists and psychiatrists who deal with youthful offenders or who seek to guide in public opinion, are largely responsible for the marked decline in moral standards among the rising generation: too often they seem ready to condone what is evil rather than condemn it". Indeed, so severe are some of the strictures, that the behaviourist now runs a risk of being rechristened a misbehaviourist.

Here, however, as mere scientists, we are concerned primarily with the *truth* of the alternative views rather than their possible consequences, though, as we shall see, the two cannot be sharply divorced. In its present phase the controversy had its origin in the sphere not of ethics but of aesthetics. Certain conclusions to which my fellow workers and I were led in the course of various experimental and statistical studies of aesthetic appreciation were subjected to drastic criticisms; the general complaint was that they were " in conflict with the subjectivist or behaviourist trends of present-day theory " (cf. this *Journal*, XIV, pp. 62f. and XV, pp. 163f). But similar objections have since been urged against the working assumptions adopted in our studies of juvenile delinquency and in our supplementary experiments on moral judgements and ethical standards. And so, in Dr. Gregson's phrase, we " come to the underlying issue of *the whole status of value in psychology* ". Much of the discussion, particularly that couched in psychological or pseudo-psychological jargon, is, it must be frankly owned, little more than what Jeremy Bentham used to call ' nonsense on stilts '. Nevertheless, it has apparently had widespread repercussions. Accordingly it would seem high time, as several contributors have recently suggested, that the arguments on both sides should be systematically re-examined, and that whatever empirical evidence is now available should be collated and re-assessed.

Let us start by inquiring in what the " present-day theory ", advocated by our critics, essentially consists. The crucial contention, we are told, is that enunciated by Mr. Paterson at the outset of the discussion, namely, that " as Ayer, Russell, and Ryle have so conclusively shown, the assumption of an ' intuitive capacity ' whereby aesthetic and ethical values are ' directly apprehended ' is an unnecessary and unverifiable postulate which modern psychology has rightly decided to drop ". Dr. Gregson agrees, but holds that Mr. Paterson " fails to press home the case he really has ". All the relevant data, he maintains, can be readily described " without any idea of value as objective "—an idea which in his opinion is " inadmissible in the context of a subjectivist ethics ". The word ' value ' need not be banned. In psychological experiments, he says, we continually find our subjects making statements which ascribe ' value ' to a thing; but such remarks must be regarded " as the way people ' verbalize experience ' ". Words like ' value ' are in effect simply " a terse way of referring to a consensus of choices at a point of time "; the psychologist himself must interpret them " behaviourally ". Thus for Dr. Gregson such utterances are just airborne noises; their meaning for the subject himself does not really count.

However, Dr. Gregson's behaviourism is admittedly somewhat diluted. Mr. Whittaker prefers it neat. With the spirited *cocasserie* that enlivens so much of his writing, he sharpens up the argument until it verges on the paradoxical; but at least we know the case we have to answer. " Biologically ", he says, " honesty simply *is* the best policy, and nothing more. The invocation of ' value ' to lend glamour to our traditional customs and conventions is simply a hang-over from a metaphysical and pre-scientific psychology, harking back to the days when psychology was a mere offshoot of theology—vestigial survivals as antiquated and as superfluous as the caudal muscles that once wagged a human tail . . . Modern behaviourism, profiting by the biological standpoint introduced by Darwin and elaborated by Freud and Watson, has shown that so-called ' ethical ' and ' aesthetic ' principles are merely the adaptive behaviour of a gregarious and highly evolved primate, which has proved requisite for the survival and progress of his race, and is learnt and relearnt by the conditioning of each individual afresh within the social group in which he is born. The man in the street ", he continues, " will no doubt persist in expressing himself in terms of hallowed notions like ' righteousness ', ' beauty ', and ' truth ', with ' happiness ' and ' bliss ' thrown in; but the psychologist, as a devotee of natural science, must scrupulously eschew all such transcendental concepts as likely to mislead him into countless fallacious conclusions ". Thus for Mr. Whittaker, it will be seen, the notion of value is a snare as well as a delusion.[1]

However, all this, as Miss Williams rightly observes, " is not so much an alternative theory as a hotch-potch of theories ". Dr. Gregson indeed has advocated at least three different interpretations corresponding to 'three levels of description'. The one point to which all our critics firmly cling is that "beauty, goodness, and the rest are *not objective but subjective*—mere reifications of each person's emotive reactions". Yet, although this conclusion is put forward as an "empirical result", it is nowhere supported by factual evidence. The knock-down argument on which they mainly rely is Occam's "principle of minimal assumptions". "With a little ingenuity all the relevant data", we are assured, "can be satisfactorily described in terms of the accepted concepts of behaviourism." Hence "the retention of specific ethical and aesthetic notions becomes superfluous, and should consequently be abandoned as involving a 'multiplication of entities beyond what is absolutely necessary'".

---

[1] It might perhaps be thought that at the very outset I should provide an explicit definition of what my colleagues and I mean when we use the word ' value ': otherwise there is certainly a danger that ' value ' may be regarded as a kind of philosopher's stone—an elusive principle, whose sole property is to turn whatever it touches into something precious. Unfortunately, as will appear in the sequel, my difficulty is that value in my view happens to be one of those primitive concepts which are not formally definable. Hence the only way to indicate how the word is used is to invite people to make an inventory of the things they value or prize. So far as I know, the earliest inquiry of this sort, carried out on a rough statistical basis, are the lists obtained by Sir Richard Livingstone from members of his lecture-classes: (*The Greek Genius*, 1912, pp. 126f.). The following are the items which he found most often mentioned, with their relative frequency in brackets: Health (54), Moral excellence or Goodness (47), Affection and Friendship (35), Intellectual excellence or knowledge of the Truth (32), Happiness or contentment (29), Artistic pleasure and Beauty (15), Physical excellence, i.e. bodily appearance, strength or skill (13), Success (13), Work (10), Money (8), Travel (8). A number of somewhat similar lists have since been obtained: (see Burt, *ap.* M. L. Reymert, *Feelings and Emotions*, 1950, pp. 544f. and refs.). The majority of the items so mentioned—e.g., money, health, strength, travel, and most of the additional items which appear in other lists, such as leisure, freedom, long life, a comfortable dwelling—are commonly regarded as *means* to further ends rather than *ends* in themselves, or, to use the customary adjectives, as ' instrumental values ' rather than ' intrinsic values '. Others, such as love or affection and happiness, I should regard as reactions to things of intrinsic value. We thus finally come down to the familiar triad of ' essential values ' that have dominated discussions of the subject from the time of the later Greek philosophers onward—namely, the Good, the True, the Beautiful, ἀγαθότης, ἀλήθεια, κάλλος (Proclus, *c.* A.D. 440). Among more recent lists probably the one most familiar to psychologists is that of Spranger—" theoretical value or truth, aesthetic value or beauty, social value or love of people, religious value or mystical union, political value or power, and economic value or what is useful " (*Types of Man*, 1928: cf. G. W. Allport, *Personality*, 1938, s.v. ' value ' and refs.). Since in the present controversy almost the only values discussed are the ethical and aesthetic, I shall confine myself mainly to these. For the history and philosophical treatment of the subject the reader may be referred to W. Stern, *Wertphilosophie* (1924), J. Laird, *The Idea of Value* (1929), and the writings of W. M. Urban, notably *Valuation: Its Nature and Laws* (1890) and *The Intelligible World: Metaphysics and Value* (1930).

The Limitations of Occam's Razor. Now I am heretic enough to plead unashamedly for what Voltaire has called *le superflu—chose très nécessaire*. It is as necessary in science as it is elsewhere. Certainly, so long as the scientist's aim is merely to *prove* things—that is, to reach verifiable conclusions which can be asserted with the highest degree of confidence—Occam's maxim, I willingly grant, provides a sound methodological canon: the fewer assumptions we include in our initial basis, the more we increase the probability of the inferences deduced from them, and the smaller the target we offer for our opponents to attack. Hence if, as Dr. Gregson and Mr. Whittaker contend, all the psychologist sets out to do is just to "construct operationally testable predictions of human behaviour—forecasts which shall have a maximum degree of accuracy", then *a priori* concepts and unproved postulates should be reduced to a minimum. But surely the scientist aspires to something more than this; he wants not only to *prove* things, but also to *understand* things. And for this second purpose a far larger array of initial concepts and postulates will nearly always be essential. In other fields, science is, and always has been, full of such speculative postulates—many, if not most of them, little more than refined versions of the familiar notions of common sense. Why then should the psychologist be forbidden the natural wish to explain and to understand as well as to prove and to predict? There is just one needful precaution. Inferences based on the two types of interpretative model—the 'economical model' and the 'explanatory model', as they may be called—must always be sharply distinguished, so that the reader should not be misled into attaching to the latter the same amount of assurance that belongs solely to the former.[1]

---

[1] Even here the distinction is not absolute, but rather one of degree. No hard and fast line can be drawn, as Dr. Gregson and others apparently suppose, between predictions which are 'operationally testable' and 'conjectural explanations' which are not: the mere fact that an assumption which is not *directly* testable nevertheless serves to increase our understanding of a particular group of phenomena is itself a partial confirmation of that assumption. I may add that the distinction between explanation and proof, for which I have pleaded in the text and in previous publications, is by no means so 'new-fangled' as Dr. Levi supposes. Every reader of detective stories is familiar with the warning that (as 'Inspector Dawson' used to say) "while the interpretation adopted by the police provides the only intelligible explanation of the facts collected, it would be hopeless on the present evidence to prove it to the satisfaction of a jury, who will never, or should never, convict unless they are convinced beyond reasonable doubt". Readers who seek a more formal account may usefully turn to what is, I believe, the most recent exposition of the notion—C. G. Hempel's opening paper on 'Explanation in Science and History' in *Frontiers of Science and Philosophy* (ed. R. G. Colodny, 1962, pp. 9f. and refs.). "Among the factors that have sustained scientific inquiry", he writes, "are two pervasive concerns: one is man's persistent desire to improve his strategic position in the world by means of *dependable methods of predicting and controlling events*; . . . there is a second—man's deep concern to *explain and understand* the phenomena it presents to him." In a paper, criticizing a recent article of mine on the 'Concept of Consciousness', Professor Keehn reminds us of Watson's pronouncement on 'the only possible programme for a natural science of psychology', namely, 'the prediction and control of behaviour' ('Consciousness and Behaviourism', (*Brit. J. Psychol.*, LIV; in the press). This formula includes only the first of the two aims mentioned by Dr. Hempel; and Dr. Gregson, we may note, pares down the Watsonian programme still further: in common with earlier psychologists of the "operational school", he restricts it to "testable predictions based on testable descriptions".

Practical Implications. The importance of this second and more comprehensive methodology becomes plain beyond all question in the case of those empirical sciences which provide the basis for practical applications. The pure or theoretical branches of science aim at the utmost degree of abstraction and precision, reducing their data so far as possible to a quantitative form, so that their deductions can be made with the accuracy which mathematical or statistical reasoning alone can guarantee: the applied or practical branches are concerned with individuals, whether persons, objects, or events, and their problems are therefore not abstract, but highly concrete. Hence the engineer, the meteorologist, the physician, and the educationist are inevitably obliged to introduce concepts and assumptions which the physicist, the physiologist, or the anthropologist would ruthlessly exclude.

Those who have criticized my "unjustified assumptions regarding the psychology of so-called values" refer chiefly to what Mr. Whittaker calls my "pronouncements as an educational, industrial, and criminal psychologist". Here, he thinks, I have "continually dragged in concepts and postulates long ago repudiated by objective psychology, which (he says) insists that all such questions can only be scientifically investigated if we treat them as we treat every other problem that arises from the conduct or the predicaments of our fellow human beings, namely, *behaviourally and statistically*". To this my answer is that, when I am studying one of my fellow human beings, my purpose, unlike that of Dr. Gregson or Professor Keehn, is not merely to "predict or control his behaviour"; it is to understand him. I am not content simply to note his "observable responses"; I want to discover his interests, his range of knowledge, his ideals and aspirations, what are his personal preferences, whether he feels happy, comfortable, or satisfied, and, in the case of younger people, what may be their innate limitations and their latent potentialities. But, like that of the Elephant's Child, my "insatiable curiosity still won't stop": I want not only to understand other people, but also to understand myself and my own conscious experiences, and (so far as I can) consciousness as such, with all that it implies.

And here, I fancy, we may trace the origin of the battle of wits that has sprung up between Mr. Whittaker and Miss Williams. For Mr. Whittaker, it would seem, as for so many of those who write and lecture on statistical psychology, all the world's a blackboard, and all the men and women merely chalk marks; for him "the end and object of psychological research", on delinquency or any other problem, is "a set of bivariate frequency diagrams showing the correlation between causal conditions on the one hand and behavioural traits on the other". For Miss Williams and for those of us who have to deal with actual persons in their everyday surroundings, their quantitative characteristics—those that can be expressed as measurements and plotted as graphs—are by no means the sole objects of our concern. We think of them as living conscious creatures like ourselves, each with his own implicit or explicit aims; and in order to discover their aims, we must first discover what things they value, and what are the priorities in their schemes of value. And whenever the case is one that called for remedial treatment, we too must have our own scheme of values to direct our efforts.[1]

However, in spite of the wide differences that separate us from our critics, there are three matters on which we all are evidently in complete accord: first, that the problems usually grouped together under the traditional labels of 'value' and 'valuing' have been sadly neglected by psychologists; secondly, that these problems urgently call for investigation by empirical methods—that is, by experiment, systematic observation, and statistical analysis; thirdly that, before any such investigations are actually planned or carried out, all the available hypotheses should be systematically re-examined, so that the main psychological issues can be clearly defined in advance.

---

[1] Let me say at once that this view is not "merely an abstract philosophical assumption which [we] see fit to adopt", as Dr. Levi and others seem to suppose. It lends itself readily to experimental and statistical investigation: see, for example, *The Young Delinquent*, 1925, pp. 405 and refs., G. W. Allport and P. E. Vernon, 'A Test for Personal Values', *J. Abnorm. and Soc. Psychol.*, XXVI, 1931, pp. 233–248, and G. W. Allport and H. Cantril, *ibid.*, XXVIII, 1933, pp. 259–273. It is much to be regretted that experimental studies of this kind have of late fallen out of favour.

Our first task must therefore be to distinguish and classify the various alternatives that have been put forward.

The Classification of Theories. Dr. Gregson's criticisms imply a dichotomous classification of relevant theories into ' objective ' and ' subjective ' respectively. He leaves us rather in the dark about which of the many possible meanings he has in mind for these familiar terms. I myself would prefer to say that the crucial distinction is between (I) those theories of value that regard it as *irreducible*, i.e., as a logically primitive concept which cannot be analysed without remainder into non-valuative characteristics, and (II) those which regard it as *reducible*.

I. My own view is that beauty is an irreducible characteristic which cannot formally be defined, but only intuitively apprehended, like the quality ' yellow ' or the relations of ' before ' and ' after '. And assuming, as in this case I presume we must, that beauty is a *predicable* quality or property (a phrase I use to cover both attributive properties and relational properties), then we may ask whether it is (i) the quality of an *independent object* (not necessarily of a physical object) or (ii) the quality of a *personal experience*. Recently, for instance, in discussing some Venetian mirrors, a colleague told me that he thought " the oval, fifteenth-century specimen far more beautiful than an irregular trapezoidal piece " (of modern workmanship); and, in reply to a further question, assured me that he considered the beauty of the older piece to form " an *objective* characteristic just like its oval shape ". Yet the " beauty " of the second movement of Beethoven's ' Emperor ' Concerto which we heard that evening he held to be an irreducible Gestalt-like quality of the auditory *experience*, and therefore " purely mental, since the actual sounds have no physical existence ": (Dr. Gregson, I take it, would agree with the latter statement, but not with the former.) In either case we may hold either (*a*) that the notion of beauty (and of value generally, including goodness) is an *empirical* concept, like green or loudness, or (*b*) that it is an *a priori* concept, much as Broad, for instance, has argued that ' cause ' and ' substance ' are ' *a priori* concepts '.

II. Theories that maintain that valuative concepts, such as beauty or goodness, can be analysed without remainder into the presence or conjunction of certain non-valuative characteristics are of many different types. The following seem the commonest, though, as we shall see later on, they are not always treated as mutually exclusive. (i) One of the oldest is the *theological* interpretation, invoked chiefly by moralists. Goodness, it has been argued, is simply what is enjoined or sanctioned by God. It is this widespread notion, I suppose, that accounts for the way in which (as Mr. Whittaker observes) so many present day behaviourists—Dr. McLeish, for example—as well as several philosophers in the past, have dismissed the suggestion that morality can be independent of human judgement on the ground that any such assumption is no more than ' an antiquated theological dogma '.[1] (ii) Secondly, there are what may be called *structural* theories, advanced chiefly by art-critics. Some writers, for example, have endeavoured to define ' beauty ' as consisting in certain structural relations exhibited by the object, which together

[1] ' Most of us would agree ', said F. P. Ramsey addressing a philosophical society in Cambridge in 1925, ' that the objectivity of good was a thing we had settled and dismissed with the existence of God; Theology and Ethics are two subjects which we realize have no real objects.' A similar view was defended by Ayer in *Language, Logic and Truth* (1936), though he has since explicitly rejected it: (see *Philosophical Essays*, 1954, Chap. X, where he starts by criticizing Ramsey's dictum).

For Mr. McLeish's argument see his recent book, *The Science of Behaviour* (1936), pp. 1f., 16, and 191). Mr. Whittaker declares that " the sheer antiquity " of my view " is sufficient to condemn it". To this we might rejoin that his own hedonic theory is just as antiquated. But the true answer is surely that of the plainspoken critic in Molière's play: " Voyons, monsieur, le temps ne fait rien à l'affaire."

constitute a coherent and harmonious whole, combining a maximum of variety with a maximum of unity. I should myself contend that, though this may be a necessary condition, it is certainly not a sufficient condition, much less that the two are identifiable. (iii) Thirdly, there are *sociological* theories. Many have maintained that what is commonly called ' right ' or ' good ' is simply a name for the rules or standards which have been enjoined by society or its representatives; and analogous ' cultural hypotheses ' have been advanced to account for our ideas of beauty. (iv) Fourthly (and frequently held in conjunction with the previous view), there is the *biological* theory, to which most present-day psychologists lend a ready support. Good conduct is defined as " adaptive behaviour calculated to promote the survival of the individual or his species ". Aesthetic qualities are interpreted in terms of sexual selection or the like. (v) Finally, there is a wide variety of *psychological* theories. (*a*) Some hold that the psychological characteristics, into which all forms of value can be analysed are essentially cognitive characteristics, e.g. that " the experience of beauty is essentially a mode of perception in which the spectator's attention is held and stimulated, and yet neither perplexed nor completely baffled, by what he sees or hears ". Similarly " moral behaviour " is described as merely " far-sighted rational behaviour ". (*b*) Some again hold that value can be equated with certain *affective* reactions, beauty, for example, with what Dr. Gregson calls ' hedonic tone ', goodness with what is termed ' true happiness '. (*c*) Others hold that value is essentially reducible to certain *emotional* reactions. According to one version—the ' expressionist ' theory—a beautiful picture is one which *expresses* both the artist's emotions and our own; according to another—the ' emotive ' theory—it is a picture which we admire or *approve of*, and similarly a good action is an action which wins our approval. (*d*) Last of all, there are theories which seek to reduce value to terms of *conative* tendencies, such as wanting, striving after, or desiring. In the last three cases beauty, it will be seen, becomes a name, not for a simple *quality* or set of simple qualities of which I am directly aware (as I am aware of pleasure, for example), but rather for a *relational property* arising out of the relation between the cognition of the physical characteristics of an object or action and my affective, emotional, or conative reaction or attitude towards it. And in each case there is a cross-classification: the defining reaction or attitude may be just one's own personal experience at the moment, and in that sense *private*, or it may be that of all normal human beings, and in that sense *public:* strictly speaking, as Broad implies, only the former group of theories should be called ' subjective ', the latter he would hold are ' objective '.[1]

The conclusion to which my own inquiries have led me is that each of the three ' essential values ' (p. 60) is an unanalysable quality, directly perceived, but existing independently of the percipient: in short, that

> Value dwells not in particular wills;
> It holds his estimate and dignity
> As well wherein 'tis precious of itself
> As in the prizer.[2]

Every one of the foregoing theories has been touched upon in the course of the present controversy. But for the most part my critics have sought to defend one of the four that I have labelled ' psychological '—sometimes, a little inconsistently, advocating two or more of the others at the same time. Accordingly in what follows I shall confine myself chiefly to these four. None of them is as new as its advocates seem to suppose. Indeed, the older historical formulation of them is usually much clearer and more intelligible than that offered by their present-day supporters. Our next step therefore must be to take each one in turn and define it as unambiguously as possible, and then briefly review the more important objections which have been urged against it. All this will inevitably involve a good many flashbacks into the past.[1]

---

[1] I base this interpretation on his remarks about *ethical* characteristics (*Five Types of Ethical Theories*, 1930, pp. 85, 260). Most psychologists seem to regard *all* ' psychological ' theories as ' subjective ' theories.

[2] *Troilus and Cressida*, II, ii, 52–5.

---

[1] The elusive terms 'objective' and 'subjective' have occurred so often in the foregoing discussion, as well as in the previous contributions to the problem, that we should perhaps pause at this point to inquire what precisely these expressions are intended to convey.

Today the word 'objective' has become a veritable shibboleth with diehards of the behaviourist school. Its popularity seems due largely to the writings of Tolman (cf. 'Psychology *versus* Immediate Experience', *Phil. Sci.*, II, 1935, pp. 354f., where it is repeatedly used to describe 'the methods of psychology as an empirical science'). Unfortunately, as a glance at the *Shorter Oxford Dictionary* will quickly reveal, the epithet may be used in at least half a dozen different senses; and none of the psychologists who borrow it deign to tell which of them they accept. Much depends on the meaning implicitly given to the noun 'object'. (i) In pre-Watsonian psychology (e.g. that of Ward), as in epistemology generally, the words 'subject' and 'object' are used as correlatives to designate the referent and the referee of an asymmetrical relation denoted by the term 'consciousness' (or its equivalents); thus "the word 'subject' (it is said) denotes the mind or ego that *has* the experience, and the word 'object' denotes whatever *is* experienced *by* a 'subject'". Whether such an 'object' is an ontological, an epistemological, or merely a psychological object are further issues into which we need not enter here. Since behaviouristic psychologists commonly avoid the word 'consciousness', and reject the whole idea of an experiencing 'mind' or 'subject', this somewhat technical interpretation is no longer open to them. (ii) Nowadays the word 'object' seems to be most frequently used as a synonym for 'material thing', so that an 'objective terminology' means what Carnap and his followers would call the 'thing-language'; and this, we are told, includes not only the names for certain things, but also the names for their observable characteristics—e.g., 'yellow' or 'hot'. On the other hand, we are expressly warned that it excludes such terms as 'temperature', 'weight', 'light-wave'. Now these are quite freely used by the behaviourist in his descriptions of stimuli, whereas he would regard attributes such as 'yellow' or 'hot' as 'secondary qualities' and therefore 'subjective'. Hence most behaviourists would presumably agree with Brunswik that the 'thing-language' as thus defined is "unsuited to supply the language of psychology" (F. Brunswik 'The Conceptual Framework of Psychology', *Int. Encycl. Unif. Sci.*, I, 1952, No. 10). (iii) Among those psychologists who accept the doctrine of physicalism the term 'objective' is often adopted as a convenient single word to designate whatever can be expressed in the language of the physical sciences, which for many physicists would mean whatever can be expressed in numerical terms or reduced to mathematical form. Since Dr. Gregson repeatedly contrasts what is 'subjective' with what is 'physical', and at his highest 'level of description' proposes to express the processes involved in terms of an algebraic equation, it would seem that this is one of the meanings he has in view. Nevertheless, few contemporary behaviourists would accept it as the only type of objectivity, since they want to include in their data such things as verbal communications and reports; and, although these certainly consist of physical movements, it would be quite impossible to express them *solely* in terms of physical movements, and still indicate their nature. (iv) Watson and other behaviourists retain the word 'subject' to denote the person upon whom an experiment or set of observations are made; and the term 'subjective' is then apparently used as a loose description for the processes or phenomena occurring within the space occupied by the subject's body—e.g. such processes as perceiving (or being stimulated), thinking (in the sense of making incipient verbal movements), 'responding emotionally' (which for them implies reactions of the subject's viscera or glands), or showing signs of pain, satisfaction, and the like. The term 'objective' consequently covers any relevant phenomenon occurring outside the subject's skin, and is virtually equivalent to 'external'—a word which is frequently substituted as a synonym. Strictly speaking therefore the colours of objects, as well as the objects themselves, should be classed as 'objective', since they are localized outside the subject's body. When the subject's physical reactions consist of movements or changes that can be observed by the psychologist, they too (somewhat inconsistently) may be termed 'objective', since they are external to the observer. This interpretation is the one adopted by Drever in his *Dictionary of Psychology* ('*objective psychology* is that type of psychology which studies phenomena that can be observed *from the outside*'). (v) Professor Broad would limit the term 'subjective' to those experiences or judgements which are 'private'—i.e., refer solely to the speaker. Thus a subjective theory of goodness is "one which would make ethical judgements consist of statements by the speaker about his own mental attitude at the time of speaking". It

II. Subjective and Non-cognitive Theories

1. *Value as based on Desire.* Let us take the last of the foregoing theories first, since that appears to be the view most frequently put forward by our critics, though usually in conjunction with one or other of the rest. "Value", says Mr. Whittaker, "is a quality or name projected on to an object to indicate the fact that it is, temporarily at least, an object of 'desire'; and 'desire' in turn is itself just another name for certain motor responses, which would lead, when finally consummated, to the acquisition of the object, as in hunger or sex. The responses are primarily reflex, but in human beings are subject to widespread conditioning. When overt, they can be observed; when merely incipient, we have to rely on verbal reports."

The earliest formulation of this theory is that of Thomas Hobbes of Malmesbury. "Whatsoever is the object of any man's appetite or desire, that is it, which he for his part calleth *good;* for these words are ever used with relation to the person that useth them, there being nothing simply or absolutely so." Hobbes goes on to argue that desire and aversion are actual "motions", being "either an endeavour toward or fromward something".[1] Later writers extended the theory to the realm of aesthetics: they would say of every form of art what Bacon says of poetry: "it submits the shows of things to the desires of the mind, whereas reason doth buckle and bow the mind to the nature of things".[2] This comes very close to the interpretation proposed by Dr. Gregson, who maintains that value is not a *quality* of some object, but implies a *relation* between the object and the subject, and then goes on to argue that "we can get round this problem" (the difficulty raised by the positivists when they tell us that any statement predicating value is a meaningless emotive noise) "by saying 'I *like* this shade of red' is behaviourally equivalent to 'this is a *nice* shade of red'". He explains that we can check or improve on the subject's 'verbal reports' about what he likes or desires by observing his overt actions, i.e. what in fact he actually chooses.

Hobbes' doctrine has gained far more support in this country than on the continent. Spinoza, it is true, has occasionally been cited as "an example of a continental philosopher who reduces value to desire or sheer conation". In Part III of his *Ethics* (ix, schol.) he declares "we do not desire or strive after a thing because we judge it to be good; we call it good because we desire it or strive after it". But this view is greatly modified by his discussion of 'the rational good' in Part IV. The supreme good, we then discover, is the intuitive understanding of the whole—the *amor intellectualis Dei.* Thus, as he says in *Letter* xxvii, "ethics rests ultimately on metaphysics".

---

follows that a "psychological theory" of value (e.g. Hume's) need not be a "subjective theory", "for value may still be public" (*Five Types of Ethical Theory*, 1930, p. 260). (vi) With these latter interpretations there is commonly combined the tacit suggestion that what are designated 'objects' or 'objective' form a class of entities or processes which, though necessarily *accessible* to human observation, and in that sense 'public', are nevertheless not *dependent for their existence* on human observation, and furthermore that they are the *only* entities or processes that have these two-fold characteristics. They thus are assumed to possess an 'absolute' reality, which cannot be validly ascribed to any other type of phenomena; all other words and phrases used in this context are mere 'verbalizations', or names which correspond to nothing that is 'real'. This often seems to be the way in which the word 'objective' is employed both by Mr. Whittaker and by Dr. Gregson (who indeed ends by mentioning the 'nominalist-realist controversy'). However, neither they nor the other behaviouristic writers who use the term mention any empirical criterion whereby the presence or absence of this implied 'reality' can be verified. It is therefore a more or less arbitrary metaphysical assumption. Let me add that, however the word may be defined, I myself should be tempted to say that, in view of our limited knowledge, any distinction between what is 'objective' and what is 'subjective' must always be a *matter of degree*, and will vary according to the context: a hard and fast dichotomy of this kind seems out of the question.

[1] *Leviathan* (1651), Pt. I, ch. vi.

[2] *Advancement of Learning* (1605), Bk. II, iv, 2.

And so we reach the familiar conclusion:

> There's no such thing as that we beauty call
> It is mere cozenage all;
> 'Tis not the meat, 'tis but the appetite
> Makes the eating a delight.
> If I fancy take
> To black and blue,
> That fancy doth it beauty make.[1]

Schopenhauer in his *Preisschrift* revived Hobbes' doctrine in its naïve form,[2] applying it to both ethical and aesthetic values; in *Die Welt als Wille und Vorstellung*, however, the theory he propounds is rather more subtle (see below pp. 72f.). But among modern adherents probably the most influential has been von Ehrenfels. " We do not " (he writes), " desire or strive after things because we perceive an intangible essence called ' worth ' in them; we ascribe ' worth ' to them because we desire them . . . Worth or value is a relation between some object and some subject, and expresses the fact either that the subject actually desires the object, or that the presence of the object conditions in the subject an experience higher in the scale of pleasure and pain than its non-existence would do."[3]

To the present-day psychologist such an account is plausible only if, like Dr. Gregson and Mr. Whittaker, he adopts a rigorously behaviouristic standpoint. He hears Tommy say he " wants to win that alley-taw ": (I borrow Mr. Whittaker's own illustration); in such a situation there is a ' subject ' (Tommy), an ' object ' (the marble), and a relation between the two, which the psychologist calls ' valuing ', and which Tommy (when he makes what Dr. Gregson calls a ' verbal report ') calls ' wanting '. For a strict behaviourist, however, the marble is and can be ' desirable ' merely in the sense that it actually is desired. For a psychologist who takes introspection seriously, however, the inferences so drawn are tantamount to arguing that the cart must pull the horse because the tail really wags the dog. " When I myself desire something ", a colleague at the Slade had just explained while discussing this very point (we happened to be talking about the purchase of some original etchings by Augustus John), " I desire it because I find it desirable in the sense of ' *worthy* to be desired '. I covet it because I consider it to have value; I do not ascribe value to it because I covet it." Each of us, I imagine, could compile a whole catalogue of things whose intrinsic value is unquestionable, which nevertheless he himself does *not* happen to want; and everyone occasionally cherishes desires which do not transmute their objects into things of value; there are evil wishes and malicious intents, as well as yearnings for what is beautiful, good, or true.

2. *Value as based on Pleasure: the Hedonist Theory*. Dr. Gregson, in keeping with many other subjectivists, takes the analysis of the concept of desire a stage further. He does not, like the out-and-out Watsonian, wash his hands entirely of the data furnished by introspection, though he insists, in the behaviouristic portions of his paper, that psychologists must regard such data merely as part of the subject's *verbal behaviour*. He then goes on to argue that, when a ' subject ' (who might, I suppose, be a poet or an art-critic) reports that he ' likes ', ' admires ', or ' wants ' such and such a thing, what he is really trying to say is, not that it has any intrinsic merit or value, but merely that " a hedonic tone is pervading the subject's experience of the object "; or, in plain language, that to admire or desire

---

[1] Sir John Suckling, *Fragmenta Aurea* (1646). Some apology is perhaps needed for the way I have interlarded my arguments with quotations; but my purpose is to show that explanations now advanced as new have already had a fairly wide vogue in the past.

[2] " Everything conformable to the striving of an individual Will ", he writes, " is called good in relation to that Will." The Will may be that of the group. Thus " an unselfish man is called good by the public because he does what serves the public or conforms with its desires ".

[3] *System der Werttheorie* (1897), Vol. I, pp. 2, 65.

a thing is simply to find a ' private joy ' in the thought or the perception of it, much as a Bentham's pig found joy in his pig-wash. The flowers of the field, outshining Solomon in all his glory, possess no genuine beauty of their own; it is only our passing pleasure that gilds the lily and lends beauty to the rose.

Hedonism of course is one of the oldest of philosophical theories, both in psychology and in ethics and aesthetics. It was advanced two thousand years ago by the Greek philosophers Aristippus and Epicurus; it was furbished up afresh in the nineteenth century by the utilitarians and their followers. " Nature ", says Bentham, " has placed man under the empire of pleasure and pain . . . His only object is to seek pleasure and avoid pain."[1] To the man of the world this has always seemed a plausible doctrine. To the mathematical psychologist it is equally attractive, because it provides a single universal currency—a hedonic scale, with a negative as well as a positive direction—along which all values can be measured.[2] Nevertheless, the whole theory has been repeatedly subjected to drastic criticism; and today, I should have thought, all who have impartially weighed the difficulties such a notion entails would be agreed that it is quite untenable. To begin with, in its commonest form it involves several flagrant inconsistencies. As ordinarily understood, it embodies two incompatible standpoints. First of all, it is held to imply, as Bentham contended, that pleasure and pain are the sources of all motivation. This the modern behaviourist would in principle accept, though he would prefer to talk of ' positive ' and ' negative reinforcement ' and leave the words ' pleasure ' and ' pain ' for the subject's verbal report: in any case he would, like Dr. McLeish, insist that the subject's actions are mechanically and inevitably determined—that he *cannot help* seeking what he calls pleasure and shunning what he calls pain. But secondly, the hedonist goes on to assure us that pleasure—" the pleasure of others as well as of ourselves "—is the only end at which we *ought* to aim. Now if we *cannot help* seeking pleasure, there is no point in telling us we *ought* to seek it. However, let us disregard the first contention, and consider only the second. There are obvious objections. To begin with, whether or not we believe we *ought* to seek it, there is little question that our most intense experiences of pleasure, like our most agonizing pains, commonly overtake us unsought and unforeseen; and as a general rule, it may fairly be said that the best way to *get* pleasure is to *forget* it. Indeed, our happiest moments are usually to be found, not when we attain our end, but during the pursuit of it. " If I held the truth in my hand ", says Emerson, " I would let it go for the joy of seeking it." And few copybook maxims are so familiar as the closing line of Stevenson's *Aes Triplex:* "To travel hopefully is better than to arrive."[3]

---

[1] *Principles of Legislation* (1789), ch. I.

[2] For a detailed outline of a ' calculus of pleasure '—' hedonimetry ', as he termed it—see F. Y. Edgeworth, *Mathematical Psychics: An Essay on the Application of Mathematics to the Moral Sciences,* (1881).

[3] According to Mr. Whittaker, however, there is a simple formula. In common with most twentieth century psychologists he believes all motivation can be expressed in terms of a single tendency. " Biologists and behaviourists, psychologists and cyberneticists ", he says, " have combined to teach us that what determines an organism's behaviour can be summed up in one elementary homeostatic law—*the recovery of equilibrium.* The environmental change that forms the ' stimulus ' disturbs the organism's existing equilibrium, and its ' reaction ' consists in a series of adaptive changes whereby the original condition of stability is restored. This ", he continues, " is all that the behaviourist means by ' need reduction ', and all that the psychiatrist implies when he says that ' the basic requirement is security ' ". Translated into hedonistic terms, " it means that the only form of pleasure (or rather satisfaction) is escape from pain (i.e. from dissatisfaction) "; translated into valuative terms, it means that a positive value must be defined as consisting merely in the absence of negative values.

This is surely to push the passion for monism and the ' parsimony of hypotheses ' to the furthest limit. So long as we confine ourselves to the *experimental* evidence (that furnished by the usual type of laboratory research on rats and dogs) the theory may sound plausible enough;

But there is a further paradox. As expounded by Bentham, the hedonist doctrine enjoins us to seek " the greatest pleasure of the greatest number ". If, however, I am to take into account the pleasure of others, I shall continually discover that this means sacrificing my own. In any case it is easy to think of situations where Bentham's basic principle, taken literally, would unquestionably lead us to choose wrongly. E.g., suppose that (i) I can either (*a*) act so that George gets five units of pleasure, and Tom, Dick and Harry none at all, or (*b*) act so that each of the four receives one unit: maximization, regardless of distribution, would require us to opt for (*a*). Yet this is plainly unjust. (ii) The amount of pleasure gained from women and wine by members of ' Abodes of Love ' is far greater, because far more intense than that gained (with a good deal of pain thrown in) by a team of pathologists investigating the causes of malaria: once again the mere maximization of the quantity of pleasure would lead us to applaud the former rather than the latter.

Mill sought to modify the crudity of the older formulation by recognizing that pleasure may differ in quality as well as in quantity. " Utilitarian writers in general ", he says, " have placed the superiority of mental pleasures over bodily pleasures, simply in their greater permanence, safety, etc., of the former—that is, in their circumstantial advantages rather than their intrinsic nature." He himself holds that " some *kinds* of pleasure are more desirable and valuable: it is better to be a human being dissatisfied than a pig satisfied, Socrates dissatisfied than a fool satisfied; and if the pig or the fool is of a different opinion, it is because they know only their side of the question ".[1] But here the word ' pleasure ' —like so many of the terms popularly proposed as synonyms or substitutes for value— becomes ambiguous. It may mean, and indeed for the psychologist it should mean, what Dr. Gregson calls ' hedonic tone '; in that case it can have no quality but its own. But the word is often used to mean pleasant states, pleasant experiences, or even pleasant objects. And these of course *may* differ widely in quality. But in this case the value depends not on the pleasure but on the quality of the experience or the object.

The objections to the identification of moral value with mere pleasure may perhaps be most succinctly summed up by applying the triple criterion which I have previously

---

but it is in the sharpest conflict with the evidence from *observation*. We have only to watch the behaviour of our tame pets in the home and wild creatures in captivity to see how false it is. The animals in the zoo, like the domestic cat or dog, live in a welfare state. Well housed, well fed, relieved of the need to hunt for prey or hide from foes, they ought, according to the theory, to be living in a paradise of stable equilibrium. In point of fact, when their environment fails to provide " a stimulus to disturb that equilibrium ", they actively try to create it. For them this comfortable Nirvana means not bliss, but boredom. The same holds good of human beings. As recent research has shown, to deprive an astronaut of all external stimulations is not to reduce him to a condition of comfortable equilibrium, but to set up a series of psychoneurotic disturbances which will compensate for the monotony that ensues. As for children and teenagers, every impartial observer must agree that what most of them seek is not security and ' stable equilibrium,' but adventure.

Mr. Whittaker sums up his theory by saying that " the object of all our efforts is to abolish effort ". I agree that what may be called ' the principle of least effort ' governs many of our compulsory and habitual activities, but it has little influence in activities accompanied by attention, and none at all in spontaneous activities. Even the self-confessed hedonist would admit that the ' pleasures of pursuit ' are more satisfying than the ' pleasures of attainment ', and the former necessarily involve effort. A few have sought to explain the paradox by arguing that the pleasures of pursuit are really pleasures of progressive attainment. But this is by no means true in every instance. We can enjoy a struggle, a fight, or a strenuous game even when we know from the outset we cannot win; and the highest and most satisfying of all pursuits—research in pure science, for example—are from their very nature unending.

I have dwelt at some length on this incidental issue, because the widespread prevalence of the simplified theory that Mr. Whittaker defends is in my view largely responsible for the way psychologists and others, who have been trained along behaviouristic or psychoanalytic lines, fail so frequently in their dealings with delinquent youths and with the moral conflicts of their patients.

[1] *Utilitarianism* (1864), pp. 11f.; Mill's italics.

suggested in the case of aesthetic values. First, the variations in the degree of positive pleasure have at best only a low correlation with variations in moral value: some of the most intense pleasures are the most ignoble. Secondly, many things which we all recognize as good are not only unaccompanied by anything that can properly be called pleasure; they may actually involve pain: "Christianity", said Goethe, "is the worship of sorrow"; and certainly the 'beatitudes' explicitly preach the value of those states or experiences which are commonly regarded as the reverse of pleasurable. Thirdly, there are many pleasures which are definitely evil, e.g., the *Schadenfreude* of the sadist. Indeed, we condemn malice as a vice, not merely in spite of, but because of the pleasure that permeates it. And finally, does not the essence of the moral problem turn on the conflict between personal pleasure and true morality? "Video meliora proboque," cries Medea, "deteriora sequor". We see and recognize the good; yet we constantly succumb to temptation.

But today, I should have thought, few psychologists and philosophers would still subscribe to such an identification in the moral realm.[1] Let us therefore turn to examine the hedonic doctrine as applied not to ethics but aesthetics. It was, we may remember, the earlier discussion of 'aesthetic factors' (this *Journal*, XIII, pp. 87–92; XIV, pp. 72–77) that gave rise to Dr. Gregson's criticisms of my views on value generally. His account receives some support from the dictionary: "Beauty", it tells us, "is that quality or assemblage of qualities which gives the eye or other senses intense pleasure."[1] Once again the theory goes back to the Greek philosophers. Plato in the *Hippias Major* says that many of his contemporaries described beauty as a kind of pleasure, and that Socrates defined it as that which gave ' profitable pleasure '. Plato himself entertained a very different view. For him the two forms of value—ethical and aesthetic—were intimately allied.[2] He would have his citizens surrounded from their earliest years with objects of beauty, so that they might eventually be led to prize the loftier beauty of holiness, goodness, and reason.

---

[1] Some comments on the foregoing arguments, sent me by Mr. Paterson, call for a brief reply. He "willingly endorses [my] contention that behaviourism and hedonism are mutually inconsistent", but maintains that "for one who like [himself] repudiates a purely behaviouristic psychology, [my] criticisms lose their force". "I should agree with Mr. Whittaker", he writes, "that 'the kind of conduct which Tommy's mother calls good is the kind of conduct that pleases her', and further that, if we ask why *that* particular type of conduct pleases her, the answer is, she believes that it will in the long run make for Tommy's happiness. But", he continues, "though pleasure and happiness are motives of behaviour, they are not *modes* of behaviour; they are *feelings*, and as such wholly subjective. Hence goodness and wickedness (and for much the same reason beauty and ugliness) are likewise subjective. Thus, I accept the conclusions which Dr. Gregson and Mr. Whittaker have drawn, though I reject their arguments."

To illustrate his point, he appends one or two examples. "Tommy sniffs. By itself the action is indifferent. But the constant sniffing displeases his mother. It therefore becomes wrong. The same holds of hitting or stealing, except that for 'mother' we must substitute 'victim'. In itself 'the primrose by a river's brim' is 'a simple primrose . . . nothing more'. But it gives pleasure to the poet, and therefore becomes beautiful. Thus, action *plus* pleasure = good behaviour; object *plus* pleasure = beautiful object."

Evidently then for Mr. Paterson there is nothing either good or bad but feeling makes it so. This view, however, plainly leads to much the same paradoxes as the version of hedonism which Mr. Whittaker supports. We cannot possibly suppose that, when Loyola "renounced the pleasures of nobility and wealth", or St. Perpetua "chose death rather than the delights of a pagan life", their motives were "some kind of pleasurable feeling". Nor can the well-known maxim "It is better to suffer wrong than to do wrong"—a principle preached and practised by Socrates as well as by Jesus—be "reduced to hedonistic terms". Where then do the fallacies lie? The first step in Mr. Paterson's argument, like that of most hedonists, involves the stock fallacy of "inverting the causal direction". Tommy's behaviour pleases his mother because she considers it good; she does not consider it good because it gives her pleasure. In the second step the issue as to whether or not Tommy's happiness is, from Tommy's point of view, just "a feeling, and as such wholly subjective", is quite irrelevant. For Tommy's *mother* her son's happiness is evidently an *objective* possibility.

Similarly in the field of aesthetics: what pleases me in a 'hedonic situation' (to borrow Mr. Paterson's phrase) is nearly always some *non-hedonic* characteristic. The beauty of the primrose is not just the temporary effect of *adding* my pleasure (or the poet's pleasure) to the flower: the beauty is, so to speak, a 'public' quality, not a mere 'private' feeling. To argue along the lines laid down by Mr. Paterson is (as Eric Gill once said of a similar theory) "to treat beauty as a kind of sauce added to an otherwise stale and unpalatable dish. A good dish needs no sauce; a good picture is pleasing because it is a good picture".

In modern times the strongest support for a hedonic theory of beauty has come mainly from psychologists. " This ", so Dr. Levi assures us, "is the theory which the whole psychology of aesthetics from Fechner to Valentine seems to confirm; beauty is just a popular name for the pleasurable hedonic tone which suffuses certain intellectual experiences; and those individuals who rank high in the measurement for the ' general factor of aesthetic capacity ' do so because by their preferences they indicate that they are the most sensitive to this type of pleasure." It is a view that was shared by a small coterie of poets, artists, and art critics towards the close of the nineteenth century;[3] but it is entirely out of keeping, not only with the attitude which most writers on aesthetics—whether subjectivists, objectivists, or expressionists—have adopted during the present century, but also with that which has dominated the general trend of art criticism throughout the past. This alternative view, with its frequent comparison of beauty and goodness, comes much nearer to the doctrine first formulated and developed by Plato and his Neo-platonist followers. We have only to recall the testimony of Dante in the *Commedia* and the *Convivio*, Spenser in his *Hymne of Heavenlie Beautie*, Shelley in his *Defence of Poetry*, Wordsworth in *The Prelude*, Ruskin in *Modern Painters*, and Bridges in his *Testament of Beauty*.

However, instead of warming up stale arguments, let us glance at one or two concrete examples. When we watch a tragedy like *Hamlet* or *King Lear*, the feeling which is uppermost is hardly to be described as pleasure. It is excitement, tension, often indeed (as Aristotle held) sheer terror and pity. The agony aroused may become so acute that children often take a lasting dislike to all forms of tragedy—as did one of my own young subjects, who remarked that she " had so hoped it would all wind up with a happy ending; why couldn't Hamlet have married Ophelia, and then forgiven the king and queen? ". Or consider a typical painting from the brush of Picasso, say the *Woman in an Armchair*. Even Dr. Gregson would hardly try to persuade us that its salient characteristic is the " pleasurable hedonic tone " with which the viewer's experience is suffused. The savage displacements of the limbs, the violent distortions of their shapes, the arbitrary conjunction of incompatible points of view—all seem deliberately designed as a vehement protest against the idea that the artist's overriding function is to give mere pleasure. The whole posture suggests a body writhing in anguish. And the intention is further underlined by the fact that the twisting of the arm over the head is the classical symbol of pain. Yet surely no serious art critic could deny the picture's claim to genuine ' aesthetic value '.

---

[1] Cassell's *English Dictionary*, s.v.; *The Oxford English Dictionary* uses the word ' delight '. Among modern philosophers Santayana is perhaps the most eminent supporter of a hedonic theory. " Beauty ", he says, " is pleasure objectified—pleasure regarded as the quality of a thing . . . Apart from ourselves there can be, in a mechanical world, no element of value whatever: in removing consciousness we remove the possibility of worth " (*The Sense of Beauty*, 1896, sect. ii, ' The Definition of Beauty ').

[2] The Greek word καλοκἀγαθία expresses the affinity between the two. Cf. *Republic*, III, 401 A–D. Rightly understood, the attack in Book X on painting and poetry as imitative is fully in keeping with the standpoint I am here maintaining.

[3] I am thinking of the ' ultra-poetical super-aesthetical ' l'art-pour-l'artism of Gautier and his followers, culminating in the manifesto appended by Pater to his *Studies of the Renaissance* (1878). It became part of the ' cult for culture ' burlesqued by Mallock in *The New Republic* and by Gilbert and Sullivan in *Patience*. But it is only fair to remember Pater's protest: " I wish they would not call me a hedonist; it produces such a bad effect on people who know no Greek."

I have found it an interesting experiment to submit to my various subjects the same painter's well-known series, *Les Deux Femmes Nues;* the first portrait—a delicate satire on the traditional conception of the nude as a delicious human shape—is the almost invariable choice among those who equate the beautiful with the pleasing; the last two versions are those selected by nearly every person, whether connoisseur or otherwise, who ranks high in our 'measurement for the general aesthetic factor'. The same holds good in other fields of contemporary art. "One of the distinctive features of modern painting, modern sculpture, modern poetry, and modern music", says James Wilson, "is its manifest reaction against the sentimentality and the saccharine sweetness that the romantic school had mistaken for art."

I recognize that some degree of pleasure (or, as I should prefer to say, satisfaction) usually accompanies the experience of objects that have worth or value. But the pleasure cannot serve as an index or criterion of the value, much less be identified with it. In the first place, there are many pleasant experiences—a hot bath, a cold shower, the smell of roasting coffee beans, the taste of mutton and mint sauce—that no one would call experiences of beauty. Perhaps then we should limit it to pleasures from the higher senses or from intellectual activities. Beauty, says Thomas Aquinas, is "id quod ipsa apprehensio placet" —that which affords pleasure through sheer disinterested contemplation. But again the definition covers far too wide a range. I was recently invited to inspect a model dairy farm: its manifest efficiency gave rise to a genuine pleasure arising from "sheer disinterested contemplation". Yet I doubt if anyone would have classed it as an object of beauty. Nature seems to have annexed the strongest pleasure, not to those things whose value is commonly described as 'intrinsic', but rather to those that have an instrumental value —particularly the personal and egoistic activities that are essential for preserving either the self or the race.[1] To suppose that the value of something good or beautiful consists solely in "the pleasure that we subjectively feel in contemplating it" is rather like supposing that the truth of an elegant mathematical demonstration consists solely in the pleasure we experience as we follow the ingenious reasoning.

3. *Value as Expressiveness.* Among those who are themselves creators of works of art—particularly the painters, musicians, and poets of the romantic and neo-romantic schools—one of the most popular notions is that beauty is a mode of self-expression. Of all the theories here discussed it is the one that had exercised the greatest influence on the development of modern art. Its supreme philosophical exponent is Schopenhauer (*Die Welt als Wille und Vorstellung*, 1819, III); but a more up-to-date version of it has since been elaborated by Benedetto Croce.

According to Croce art is "a language expressive of feeling". "A picture, a poem, or a sonata is beautiful to the man whose feelings are expressed in it." When we look at the picture or listen to the poem or the sonata, we recognize in it emotions which we attribute to a fellow-creature like ourselves who expressed what he felt by painting the picture, or by writing the poem or the sonata. The " true function " of a work is thus " to express emotion ". But there is no need for it to represent natural objects in order to have something that will serve as intermediary vehicles for the emotion: the emotion can, so to speak, be delivered direct from the manufacturer to the customer. Nevertheless, though what we really contemplate is volitional or passionate, the aesthetic experience as such is ' theoretical ', i.e., cognitive rather than emotive: the *esse* of beauty is *cognosci*. Since the artist's own experience is always individual, it is incapable of description in scientific terms; hence we recognize it, not by means of our reason or our understanding, but by a kind of empathy, or, as Croce prefers to say, ' intuition '.[1]

---

[1] Herbert Spencer, and the later psychologists of the utilitarian school who adopted his evolutionary theories, maintained that *all* values are instrumental values or the result of associations with such values. In origin they owe what we now call their value to the way they have facilitated, and often still facilitate, the adjustment of the individual and the group to the environment—particularly (in the later stages) to the social environment, and so make not only for the survival but also for the evolution of the species. With later writers this biological interpretation was frequently combined with Schiller's well known 'play-theory' and with Freud's so-called 'pleasure principle'. In this way, it was held, both ethical and aesthetic judgements could readily be derived from non-ethical and non-aesthetic principles. As an account of the origin of such judgements the theory no doubt embodies an important element of truth. But the inferences commonly drawn rest on one of the commonest fallacies, namely, the assumption that, in describing the *origin* of a thing, we have also explained, or explained away, its *validity*. Because chemistry has developed out of alchemy and astronomy out of astrology it does not follow that the validity of the speculations of the modern chemist and the modern astronomer can therefore be rejected without further examination. Today such evolutionary interpretations still seem popular with anthropologists and sociologists, but are relatively rare in psychological writings. They have not been advanced by any of my own critics, and so call for no detailed discussion here.

Among psychologists, save for the observations and inferences of certain leading psychoanalysts, this view has been unduly neglected.[2] As expounded by psychoanalytical writers the expressionist theory has a double application ; it maintains, first, that " in creating a work of art—whether picture, poem, or musical composition, the *artist* is primarily expressing his own emotions, often quite unconsciously ", and secondly, that " in looking at the painting, reading the poem or hearing the music, the *spectator or listener* interprets it, again to a large extent unconsciously, as an expression of *his* emotions ".

(i) The first part of the theory is commonly supported by a variety of clinical observations. Time after time, we are told, the psychoanalyst finds that what a patient writes or draws or paints, both by its content and by its style, throws a flood of light on his conscious and unconscious interests and motives; and this, it is argued, implies that " his interests and motives, and indeed his whole history and social background *find a condensed expression* in his verbal or pictorial productions ". Nevertheless, such observations scarcely justify us in erecting this useful psychodiagnostic principle into a theory of normal art-production. After all in this broad sense almost everything we do or say is a form of self expression. Scientists, like Galileo and Newton, no doubt are partly influenced in their choice of topics and treatment by the social conditions of their time and the circumstances of their own past life and individual temperament; but in the main their choice expresses " their current feelings, interests and problems ": nevertheless that does not warrant our declaring that " the function of a work of science is to express emotion ". In spite of Dr. Rosenthal's illuminating ' depth-analysis ' of Turner's Alpine paintings, an untravelled analyst would, I am convinced, learn far more from them about the Alps and their beauty than about Turner and his complexes. The emotional characteristics of the artist (other than his aesthetic experiences) are but one incidental factor in the work that he produces.[3]

---

[1] B. Croce, *Estetica* (1902), esp. chaps. XIII–XV. Since the first function of language (as the modern theory of linguistics assures us) is communication, i.e. the transmission of experience, it would seem that in contemporary terms Croce's theory is really a theory of art as ' communication ', of which ' expression ' is merely a limiting case. When a man gives vent to his feelings by means of words, whether in prose or in verse, he nearly always assumes that someone besides himself will hear or read what he has uttered; and few paintings are pictorial soliloquies.

[2] The psychoanalytic theories started with Freud's famous study of a painting by Leonardo, Ernest Jones' study of *Hamlet*, and Karl Jaspers' study of Van Gogh. In what follows I am much indebted to Ernst Kris, *Psychoanalytic Explorations of Artistic Vision and Hearing* (1953), and above all to E. H. Gombrich, *Art and Illusion* (1960), although I myself cannot agree with their own conclusions.

[3] According to the psychoanalyst the painter ' interprets ' the landscape before him much as a patient interprets a Rorschach inkblot, and in both cases, it is argued, the interpretations " tell us far more about the interpreter than they do about the object interpreted ". The device of using blots and splashes to stimulate the imagination is at least as old as Leonardo. In the 18th century Alexander Cozens wrote a manual on *A New Method of Assisting the Invention in Drawing Original Compositions in Landscape* (1785); but the landscapes which the author himself claims to have seen in the blots he reproduces are remarkably like the type of landscape then made fashionable by Claude

(ii) What the *spectator* sees in a picture, we are told, is obviously influenced by his own emotional attitudes and predispositions. But, as Professor Gombrich has so clearly shown, the spectator's interpretation depends much more on the mental stereotypes—the 'apperceptive schemata', I am tempted to call them—which he automatically applies when he recognizes what the picture is intended to represent; and these schemata tacitly presuppose the techniques and the conventions regularly used in paintings with which he is already familiar.[1] To those accustomed to the style of Cimabue and his school the 'modern manner' of Giotto (so Vasari tells us) seemed startlingly natural.

Lorraine. On the other hand, in the following century Justinus Kerner (the physician, poet, and spiritualist, who today is chiefly remembered for the famous student song "Wohlauf noch getrunken . . . Juvivallera") used inkblots (*Klecksen*) on folded paper as a stimulus for his poems (*Klekso-graphien*, 1857). Being himself an ardent spiritualist and an intimate acquaintance of Frederika Hauffer, 'Die Seherin von Prevorst', what he usually saw in the blots were demons and the spectres of departed friends.

Turner used to give a paint-box to small children and watch what novel shapes or colour harmonies would appear in their random daubs. Klee and Picasso formed the habit of "doodling to see what would turn up". Galton relates that, having heard of Turner's experiments, he himself deliberately practised "the automatic construction of figures by allowing my hand to scribble at will while I gave my best attention to what was being said by others"; he notes that "it was always a surprise" to see the fantastic results that eventually emerged.

In what follows my intention is not to decry the method as a useful instrument for exploring individual differences. Indeed I myself long ago used ink-blots and pictures, ambiguous and otherwise, as a method of probing emotional tendencies and complexes: (cf. *Mental Differences between the Sexes*, J. Exp. Pedag. I, 1912, pp. 273f. 355f.; also Amya Sen, *A Statistical Study of the Rorschach Test*, this *Journal*, III, 1950, pp. 21f. and refs., and *id.*, *A Preliminary Study of the Thematic Apperception Test*, *ibid.*, VI, 1953, pp. 91f.). The notion that gazing at a set of pictures may precipitate a revealing 'projection' of the spectator's deeper emotions is by no means a new discovery. In Shakespeare's *Rape of Lucrece* the victim's eyes light on "a piece of skilful painting", depicting scenes from the siege of Troy; "much imaginary work was there . . . A hand, a foot, a head stood for the whole to be imagined . . . And so Lucrece, set a-work, sad tales doth tell to pencill'd pensiveness and colour's sorrow. She lends them words, and she their looks doth borrow". And who can forget Polonius's apperceptions of the cloud, as a camel, a weasel, or a whale, or Hamlet's confrontation of his mother? "Look here, upon this picture, and on this."

[1] As we have learnt from studying the development of drawing ability in children, the youthful draughtsman's first efforts begin with what are in effect descriptive schemata or symbols representing, not what he sees, but his conceptions of what he sees. And when he strives after greater realism, he learns far more by copying the drawings in his illustrated story-books than by trying to reproduce the objects he actually beholds. When later on he attempts to sketch from nature, it is still the schemata that fit his own apperceptive images and his own habitual techniques that stand out and catch his attention: (see Burt, *Mental and Scholastic Tests*, pp. 417–429 and figs. 53–64). And the same holds good (as Nietzsche ironically remarks) of the professional artist.

Treu der Natur und ganz! Wie fängt er's an?
Er malt zuletzt davon, was ihm gefällt.
Und was gefällt ihm? Was er malen kann!

The catalogues of the larger art libraries include the titles of innumerable drawing books and manuals of painting going back to the 16th century and earlier. Even the most original artist has always begun by appropriating from his predecessors a kind of mental pattern book containing useful pictorial schemes, and then refined upon them. These he uses to classify, according to *genus* and *species*, what he observes; he then modifies, corrects, and adjusts the scheme to represent the *individuality* of the objects and the scene before him. Once more the basic tendency is neatly described by a poet, W. H. Auden:

An abstract model of events
Derived from dead experiments
Becomes, though still particular,
An algebraic formula.

Credette Cimabue nella pintura
Tener lo campo; ed ora ha Giotto il grido,

says Dante. We ourselves, with Raphael's Madonnas in our minds, find Giotto's 'manner' archaic and stiff. If Gainsborough's pictogram for a tree corresponds with ours, we think it marvellously realistic; and by contrast Constable's 'snow' appears a very unconvincing expedient for suggesting the glitter of raindrops on the leaves of British elms. Then, we gradually learn to accept and to admire Constable's conventions. Later still, as we turn from Constable to the landscapes of Corot, Monet, Cézanne, Utrillo, or Hokusai, we have in each case to increase our apperceptive resources and learn a new pictorial vocabulary and grammar. Most of us who in our youth were inured to the 'kanons' or norms laid down by Pheidias and Praxiteles have needed a good deal of rehabituation before we could appreciate the genuine beauty of the best Egyptian sculpture or the statuary of Henry Moore. Thus, if art is 'expressive', what it expresses are the artist's schemes or conceptions quite as much as his emotions or feelings; and yet, at its best it can, by means of these conceptions and emotions, effect a unique poetic evocation of what is (I maintain) objectively beautiful.

The experienced artist himself is well aware of the extent to which he can rely on the interpretations of the contemporary spectator. Leonardo, the pioneer of naturalistic painting, knew how much a picture could gain in realism when some of the realistic details were left out or reduced to the barest hints. Even the most casual visitor to our art galleries cannot fail to note how, in Hals' *Laughing Cavalier* or Sargent's *Carmencita*, what seem, while he stands near enough to 'smell the paint', to be simply slapdash brush-strokes, miraculously resolve themselves at a few yards' distance into a rich embroidery or an intricate lace. And today, the *sfumato*, the *bravura*, or the *sprezzatura*, which were once the hall marks of skilful painting, have turned into mere technical tricks which every twentieth-century caricaturist and advertisement designer will exploit *ad nauseam*.[1]

I should make much the same reply to the expressionist theories of music and architecture developed still more recently both by writers on aesthetics and by practitioners themselves. But a detailed discussion would prove rather more technical and require more space than is practicable here.[1]

---

[1] The way in which what we mistakenly suppose to be direct perception proceeds by filling out fragmentary clues is admirably exemplified by our experiences at the theatre. Seen from the back of the stalls, the heroine looks exquisitely beautiful; inspect her through your opera-glasses, and too often the illusion is completely shattered.

The ambiguities to which these devices may give rise when pushed to their furthest limits, have of late been ingeniously developed by the Cubists, the German Expressionists, and the various other schools into which the modernist movement has eventually split—most of all perhaps by the socalled 'psychoanalytical art' conceived by that *enfant terrible* of Surrealism, Salvador Dali. He begins, like Picasso, by breaking up the basic unit of our daily practical world—the distinguishable 'object' or 'thing', thus splitting the humdrum atom, as it were—and leaves his explosive extravaganza without any single eye-catching element to focus the attention. All natural shapes are reduced to their simplest geometrical forms, and squashed flat against the picture-plane and the various disconnected fragments are thrown together with all the inconsequence of a dream. The spectator—when he gets over the first visual shock—is thus left to tease himself by apperceiving or interpreting what he sees first in one way and then in another; and so in the end may (or may not) find that, in some subtle or sophisticated fashion, they express certain inexpressible or unsuspected feelings of his own. As he stares at these 'cryptograms on canvas', the critic who has been brought up on more orthodox styles may well be tempted to echo the Devil in Kipling's *Conundrum*:

The Devil whoops, as he whoop'd of old:
" It's clever! But is it Art? "

I fully agree with Dr. Levi in deploring the way psychologists have neglected 'the rich field of modern art and music.' The problems are too complex to discuss in detail here. I can only hope that the brief remarks that follow will suffice to dispose of the argument that " subjectivity of modern art is enough to show that artistic values are essentially subjective ".

The fatal objection to a purely expressionist theory of beauty lies in its inapplicability to the contemplation of Nature. As an explanation of our reactions to the work of another human being, blessed with feelings and emotions similar to our own, the theory may seem plausible enough; but it cannot account for our discernment of beauty in a flower, a landscape, or a night of stars. Partly no doubt for this reason Croce, like Kant and Hegel before him, deliberately excludes Nature and natural objects. To compare the Venus of Milo with a landscape by calling both beautiful, he says, would be "blasphemy". Yet could she not be compared with a beautiful woman? One or two writers, it is true, have maintained that "we only find Nature beautiful in so far as she reflects our own emotions". But such an argument is tantamount to erecting what Ruskin[1] has taught us to call 'the pathetic fallacy' into a universal principle. Moreover, it is an inversion of the expressionist theory in its original form. According to that the beholder sympathizes with the feelings latent in what he beholds; and now we are required to assume that what he beholds in Nature appears to sympathize with the feelings latent in the beholder.

---

[1] Roughly synchronizing with the various post-impressionist movements in painting, i.e. just before the first world war, there was a determined attempt to win a greater freedom in music by breaking up the units of the traditional style and abandoning the familiar organization into readily recognizable keys, rhythmic patterns, melodic outlines, and harmonic sequences. Bars and key-signatures were deleted; the rules of consonance and harmonic progression defiantly violated; and a 12-note scale introduced to obviate tonality. The ideals of the romantic school a century before were thus carried to the furthest limit: the aim of music, as of painting, was to express the 'inner experiences' of the artist. Scriabin, for example, in his later works (e.g., the *Poem of Ecstasy*) evolved a 'mystical chord' of six superposed 4ths (four perfect, two augmented), and even projected a 'Mystery', in which dance, speech, music, colour, and perfume were to be combined; the results were to embody the 'theosophical illumination' derived apparently quite as much from the writings of Madame Blavatsky as from first hand 'ecstasy'. The leader of the 'expressionist' school was Schönberg—a friend of Kadinsky, the founder of expressionism in painting and later (with Klee) professor at the Bauhaus. Schönberg was himself a painter in the same style.

The first impression conveyed by the new music was one of complete individualistic anarchy. But closer study showed that the rupture with tradition was far from complete. Scriabin himself complained that he could never escape the implications of the dominant 7th. Schönberg supplied a textbook, beginning with a masterly analysis of classical harmony, and then attempting a theoretical basis for the new style. Rigorous rules were prescribed, e.g., that no note can be repeated until all the others in the chromatic scale have been sounded. And, when all else failed, 'the unconscious' was freely invoked to lend an artistic verisimilitude to an otherwise bald and unconvincing explanation.

Thus, here as elsewhere, *order insists on reappearing*. With patience and concentrated study the sympathetic listener gradually feels his way to the schemata implicit in each composer's works (often traceable to earlier sources, notably Liszt, Wagner, and Debussy); and with these additions to his apperceptive resources, he becomes able to appreciate the almost neurotic sense of tension, conflict, and revolt, which so much of the earlier music expresses, and to enjoy the novel possibilities opened up by these daring experiments, ready for later composers in the modern style.

Somewhat similar developments took place in architecture. But, partly as a result of the general impoverishment that followed the first world war, especially in Germany, the *emotional* expressionism of the pre-war period gave way to a *functional* expressionism, in which the aim was to express, not the artist's or the community's conscious or unconscious emotions, but rather the most economical adaptation of the work of art to its specific purpose. The highest type of value was a utilitarian or instrumental value rather than a psychological. Artistic creation itself was now subjected to Occam's razor. Art and technology—particularly structural engineering—became allies rather than enemies. And here to a large extent the severely abstract tendencies of Bauhaus architecture ran clearly parallel with the desire for 'abstraction' instead of 'representation' in painting and music.

Today, however, in each of the fine arts we may, I fancy, detect a reaction and an incipient return to traditional principles in a more or less modernized or sophisticated guise. Klee in his latest works, as his posthumous *Bildnerisches Denken* (1956) explains, came to regard his art as "an attempt to communicate a world as it might be or is in process of becoming". Klee and Kandinsky were musicians, and gave *musical* titles to their works, for music is abstract. But in the past the most influential painters—Michael Angelo, Blake, Turner, Rosetti, and the best known Chinese artists—were more often pre-eminent as *poets*. So, too, in a minor way are several of our contemporary innovators. Indeed the prolonged taboo on 'literary associations' is at last being lifted once again; and, while I am writing this note, there is in London an exhibition of work by Kurt Schwitters which he entitles 'The Visual Images of a Poet',—hailed by the critics as "typical of the new revival—a gradual reappearance of the concrete in place of what is purely abstract". Thus the order emphasized by the abstract painters is revealed once again in the concrete.

Bosanquet, however—the only philosopher of eminence who has attempted such an extension—has given the theory a slightly different turn. He proposes to define beauty, both in art and nature, as "that which has individual expressiveness".[2] But from the explanation that ensues it seems that by *expressing* our emotions he really means *stimulating* our emotions. And then we immediately inquire why some objects or scenes should stimulate our emotions in such a way as to excite aesthetic appreciation while others do not? The answer presumably is contained in the phrase "conditions for general or abstract expressiveness"; and when we ask what precisely this includes, we find after all that it covers almost exactly those qualities or conditions which believers in objective beauty ascribe to beautiful objects, and that these are 'directly apprehended'.

4. *The Emotive Theory of Value*. There is, however, yet another type of 'affective subjectivism' which has recently come back into favour. Hume in his *Treatise of Human Nature* (Book II 'Of Morals', 1739) endeavoured to show that "moral distinctions cannot be derived from reason". If, for example, we argue: "Working at a useful occupation is a method of increasing useful commodities; increasing useful commodities is a result which will bring pleasure to a large number of people; therefore I ought to work at a useful occupation", then (he maintains) "this *ought* expresses a new relation or affirmation which cannot be deduced from propositions which merely contain *is*." Hume therefore concludes that assertions which contain *ought* or *ought not* do not state reasoned inferences from facts—though reasoning may be needed for a preliminary analysis of the facts; they merely express "emotions of approbation or disapprobation". And he adopts much the same theory to explain statements about what is or is not beautiful. But at this point we should perhaps recall Wordsworth's well known comment—that "Hume was the worst critic that Scotland produced".[3]

Hume's theory has since been revived and popularized by Professor Ayer. "Aesthetic words, such as 'beautiful' or 'hideous'", he writes (*Language, Truth and Logic*, 1936, pp. 109f.) "are employed, as ethical words are employed, not to make statements of fact, but to express certain feelings and evoke a certain response." And once again we ask, what can these feelings and responses be if they are not just feelings or responses of pleasure? In the case of ethics, we are told, there is "a special sort of *moral* approval or disapproval which is the feeling that is being expressed". In the case of aesthetics it is presumably a special sort of *aesthetic* 'approval or disapproval'.

---

[1] *Modern Painters*, III, iv, chap. xii (The Pathetic Fallacy).

[2] 'On the Nature of Aesthetic Emotion', *Mind*, III, no. 10.

[3] What Hume is really attacking are theories like those of Clarke and Locke (Essay, III, xi, IV, iii), who had sought to refute Hobbes by showing that moral duties were capable of rational demonstration. He himself is to some extent influenced by both Butler and Hutcheson (see below pp. 83f.), and is in effect attempting to reconcile the notion of a 'moral sense' (see title to Sect. ii of Bk. III) with a utilitarian theory not unlike that of Bentham and Mill. His final views are most clearly expressed in his *Enquiry concerning the Principles of Morals* (1751), which he described as "of all my writings incomparably the best". Aesthetic values, he tells us, are on much the same footing as moral values. In the *Treatise* (Bk. II, Pt. 1, sect. viii), it is true, he at first says "beauty, like wit, cannot be defined, but is discerned only by taste". But elsewhere he argues that it is a kind of implicit utility, and so produces pleasure in the beholder. "Most works of art are esteem'd beautiful in proportion to their fitness for the use of man." In his essay *Of the Standard of Tast* he says of aesthetic appreciation, as he has done of moral judgement, that "the principles of taste are universal, and nearly if not entirely the same in all men, though few are qualified to give judgement on any work of art".

Now as a psychologist I find it difficult to regard approval or disapproval as a feeling or an emotion. To approve, so the *Oxford Dictionary* tells us, is "to pronounce or consider good"; and to do that is surely to make or utter a judgement, not to "express a feeling". And this is the view of so careful a psychologist as McDougall. "Approval and disapproval", he observes, "have been treated by Westermark and others as emotions. But to describe them as emotions is to perpetuate the chaos of psychological terminology. They are not emotions but judgements." Like other judgements, they may be partly determined by emotion; but they may also be unemotional and purely intellectual conclusions, "made in logical accordance with previously adopted principles".[1]

However, let us waive these points of psychological nomenclature, and consider what precisely the theory implies. And at the outset we naturally inquire—whose approval or disapproval? In the case of ethics there is a pretty wide choice: it may be the approval (i) of the speaker himself (Ayer, for example, commonly talks in the first person singular), (ii) of 'us', i.e., the particular philosophical group which the speaker represents (usually designated without clear specification by the first person plural), (iii) of all mankind, except perhaps for a few abnormal deviants, such as defectives, lunatics, and criminals, (iv) of all persons of sound moral judgement, (v) of the majority of those in our own social group, (vi) of those who in virtue of their class or office can authoritatively decide what conduct should receive approval—the priests, the prophets, and the parsons, the legislators and the judges, the powerful and the privileged, or possibly those who hand on to each succeeding generation the accumulated results of past traditions—our elders and our parents, and finally (vii) of God Himself. Roughly speaking, the interpretations might be grouped together as the aristocratic, the democratic, the oligarchic, and the monarchic or theological. Each of these alternatives has had its advocate in the past; but there is no need to examine all of them in detail here. None of my critics takes his stand on the last two alternatives. Several, however, appear to favour what may be called the democratic alternative; and, since no one today would hold that all mankind except the criminals, lunatics, and defectives are agreed in what they approve, we should perhaps accept the fifth alternative, namely, that value is that which can claim the approval of a majority in our own social group.[2] But this would seem to reduce ethics, aesthetics, and logic to a branch of statistical psychology based on counting votes—a kind of criticism by ballot box. And, since the standards of the majority are admittedly cruder than those of the philosophic few, we should be forced to agree with Bentham that, "prejudice apart, the game of pushpin is of equal value with the arts of poetry and music", and if we lived in a community where the masses revelled in watching a witch or a heretic suffering agonies at the stake, we should then be obliged to value what so plainly gained the general approval.

We thus seem to be thrown back on one or other of the first four assumptions. Now if we say that the type of approval we have in mind is that of 'persons of sound moral judgement', we shall be guilty of a *petitio principii*: for we shall then very naturally be asked, how are we to determine which particular persons possess a sound judgement, and the answer can only be those whose judgements enjoy the approval of ourselves or of their fellows. A factorist might perhaps suggest a correlational approach: we could invite a large and random sample of people to rank in order of goodness a series of typical actions from the most virtuous to the most vicious, correlate their rankings, and, if the correlations were predominantly positive, extract a general factor, namely, the average order of the whole set calculated by weighting the individual orders. We could then define 'a person with a sound moral judgement' as one whose ranking showed a sufficiently high correlation with the ideal order obtained for this factor. But that of itself could not serve to buttress up the subjective theory: indeed, if the correlations proved to be mainly positive (as in fact they do), the result would tend strongly to support the theory of objectivity. However, neither Ayer, nor any of the philosophers or psychologists who have followed him, have proposed a factorial solution: on the contrary, they seem rather to suppose that the correlations would for the most part approximate to zero, so that on their view the method would be inapplicable.

---

[1] *Introduction to Social Psychology* (1946), p. 124.

[2] This seems to be the form in which the theory is adopted by most behaviourists. Skinner, for example, writes "'You ought to love your neighbour' may be converted into the two statements: (1) 'Loving your neighbour is approved by the group of which you are a member'; (2) 'The approval of your fellow-men is *positively reinforcing to you*'. Both these statements may be demonstrated scientifically." (*Science and Human Behaviour*, 1956, p. 429; slightly abridged.)

Thus their final answer to our preliminary inquiry must be that they are talking of the judgements of each individual considered separately, with their own taken as typical. Ayer's actual wording is: " If *I* say to someone, ' You acted wrongly . . .', *I* am not making any statement; . . . *I* am simply evincing *my* moral disapproval." And presumably everyone is supposed to talk in the same personal way. It should be added that what he says of ethical judgements is intended to hold equally of aesthetic judgements, since, " in their case ", he says, " the same argument also applies ". What then is the argument?

If we turn to the original account of the theory as outlined by Ayer in the first edition of *Language, Truth and Logic* (to which Dr. Levi refers us), we learn that Ayer's main reason for putting forward an ' emotive theory of value ' was that the objectivist doctrine " presents an insuperable objection to our radical empiricist theory ". Consequently he felt it was his " business to give an account of ' judgements of value ' which would be both satisfactory in itself and consistent with our general empirical principles ". For the rest his own defence of the view is relatively brief.

So far as moral judgements are concerned the most detailed exposition of Ayer's view is that of C. L. Stevenson in his book *Ethics and Language* (1944). The general line he follows resembles that taken by Dr. Gregson and Mr. Whittaker; but his arguments are fuller and more systematic. He disowns any claim to *prove* the view that he advocates. He is content, he says, to explain it, and demonstrate that it is more plausible than any other. He begins, as Ayer did, by contending that, when people pronounce ethical judgements, they do not really mean to make assertions about definite facts. What they are doing is to communicate the emotional attitudes or feelings which they themselves experience. In a police court I could offer factual evidence to prove that John Doe stole my watch; but I could not offer factual evidence to prove that he acted wrongly when he did so. (The utilitarians who support a naturalistic view of ethical judgements would maintain that they *could* offer such evidence, but I should readily agree that their alleged demonstrations were invalid). Since therefore the assertion that stealing is wrong is not a proposition that I can verify by observational evidence or by any testable criterion, it follows, so Stevenson and his fellow positivists contend, that the assertion is meaningless. It expresses not an idea or a conclusion, but merely an emotion or feeling. This is in keeping with Professor Ayer's contention: " To say, ' You acted wrongly when you stole that money ' does not state anything more than the concrete fact alleged, namely, ' You stole that money '. The words ' acted wrongly ' are equivalent to an exclamation mark: it is as if I had said ' You stole that money ' in a peculiar tone of horror." And doubtless the ' peculiar tone of horror ' is intended at the same time to dissuade the pilferer from repeating his offence.[1] This interpretation, it is argued, furnishes an entirely adequate explanation of what are miscalled ' moral judgements '; hence there is no need to invoke any further assumption to the effect he was not to blame or even from a 'true criminosis' (whatever that may mean)—anything rather than recognize that he is morally responsible. Nearly every psychologist seems agreed that the State should not administer punishment by way of *retribution* for the crime, i.e., to vindicate moreal standards, first because "the members of the State are rarely unanimous about the wrongfulness of particular actions", and secondly because both psychoanalysis and behaviourism postulate a "strict determinism", which is wholly incompatible with the conclusion that the wrongful act might result from free and voluntary choice. Deterrent punishment and reformative punishment they are ready to allow in theory, since these are consistent with the principles of 'conditioning' and 'reinforcement'. Yet in practice the main tendency is to deprecate all form of punishment in favour of quasi-psychiatric 'treatment' (which in point of fact is seldom attempted even in the rare cases where it is manifestly needed).[1]

---

[1] Cf. Ayer, *Language, Truth and Logic*, p. 107.

The most recent as well as the most outspoken statement of the behaviourist position is that contained in Dr. John McLeish's recent book, *The Science of Behaviour* (1963). It has already earned high praise from at least one eminent professor working in the field of personality. McLeish's conception of psychology, like that of Watson himself, is based essentially on the Pavlovian approach. "Pavlov's theory", we are told, "is a model theory for the science of behaviour." In place of the older dualism, which recognized both conscious processes and material processes, Pavlov, we are told, has taught us to substitute a materialistic 'monism'; what is subjective must be ruthlessly rejected in favour of what is purely objective. And in this way psychology is at last "freed from the Procrustean framework of introspection and theology". The cardinal error of all "theological explanations of behaviour", it is argued, was to "posit a freedom of choice on the part of the behaving subject." By making "the principle of determinism basic" behaviourism has succeeded in relieving us of all this "arbitrary hocus pocus".

Chapter II of Dr. McLeish's book works out the corollaries of this doctrine for the problems of 'sin, crime and criminal responsibility'. Since "the whole theory and practice of our criminal law is built up on the assumption that the criminal *intends* his action," and "intention depends on free choice", our current attitude, both popular and legal, must be drastically changed. Dr. McLeish admits that "more recently the quantum physicists have suggested that science must discard the notion of cause"; but (so he assures us) "this suggestion is not regarded favourably by scientists"—an amazing statement in view of the fact that all except one or two unorthodox critics now accept "the principle of indeterminacy". He recognizes too that there are still a number of old-fashioned moralists who "assert that this [behaviourist] theory is destructive of law and morals;" and as a result "psychologists are now often accused of weakening the moral fibre of present day society". Nevertheless, in psychology as elsewhere the causal law is inexorable. The archaic "religious assumptions" on which the popular notions of mind, will, purpose, moral values, and the like have been based "are excluded": they "merely multiply entities which are unnecessary in the explanation of behaviour". And thus our final conclusion must be that "the whole idea of sin and delinquency is alien to psychology".

Whether there may be any actual facts or data to support or explain the popular notions of 'intention', 'consciousness', 'purpose', 'value', etc., Dr. McLeish does not stay to inquire. It is curious that a writer who insists so strongly on 'objective' methods should decide these particular issues simply on grounds that are flagrantly 'subjective'—namely, his own personal preference for two 'basic assumptions', first, 'the canon of simplicity', which enables him to dismiss in advance anything that does not fit his own simple-minded theory as 'wholly superfluous', and secondly the principle of 'causation', which is itself (as Russell and others have repeatedly pointed out) "one of the most subjective notions that have ever found their way into science". It may perhaps seem wrong to blame Dr. McLeish for these rather dogmatic pronouncements, since according to his hypothesis his own behaviour must also have been 'automatically determined'. Yet, strange to say, having spent so much time on insisting that the notion of 'intention' must be eliminated as contrary to universal determinism, he ends on the very last page of his book by explaining what has been "the intention of the author". If then authors—or at any rate this author—can have 'intentions', why not criminals?

---

[1] See P. Callard, 'Punishment by the State', *Brit. J. Delinq.*, X, 1959, pp. 36f. For the term 'criminosis' see *ibid.*, X, 1960, p. 175.

Mr. Whittaker's approach (pp. 48f.) is avowedly the same as that of Dr. McLeish. On the question of criminal responsibility and kindred problems the views adopted in *The Young Delinquent*, and by "those who still follow its pre-behaviouristic principles, are" he tells us, "sadly behind the times". However his attack on the conclusions set forth in that volume is evidently developed, like that of Dr. McLeish, purely from the abstract standpoint of the psychological theorist. May I therefore, on the basis of a fairly long experience both of school children and of work in juvenile courts, try to answer his criticisms from the practical standpoint? Dr. Whittaker argues that those who accept the principles assumed in the chapters he criticizes "are seeking to impose an immutable scheme of ethical commandments or norms dating back to Sinai".[1] This seems to be the same type of misunderstanding that he shows in discussing so-called 'aesthetic norms'. In point of fact I can find no such doctrine in the book he has attacked; and in spite of all his protests it must surely be acknowledged that its two main conclusions have since been amply substantiated by actual events. The first and most important result was the demonstration of the 'plurality of causes', (p. 590). This, strange to say, is still ignored by the vast majority of current researches, which endeavour to establish some single, highly vague and generalized factor as the essential cause. The choice of the factor varies from one investigation to another; but all have one thing in common: they scrupulously avoid any ethical reference, and tend to substitute a pseudo-medical or quasi-sociological terminology for plain moral concepts. The second conclusion was itself twofold, namely, that "*material* conditions, such as poverty are far less important than *moral* conditions" (p. 607). When the book was published, nearly every criminologist of the day laid the greatest stress on material conditions; by abolishing poverty, it was argued, the welfare state, when at last it arrived, would go far towards abolishing delinquency and crime; and Burt's conclusion was denounced by contemporary reviewers as "contrary to all experience".[2] He on the other hand laid chief emphasis on psychological factors, and most of all on the lack of any guiding purpose in life.

---

[1] To avoid this type of criticism may I say that those of us who took part in the investigations fully recognized that one of the main difficulties confronting us was precisely the way in which the nominal code was out of step with the rapidly changing attitude of the times—particularly in regard to such matters as parental authority, sexual ethics, homosexuality, suicide, and the like? However, as Burt warned us in advance, "the psychologist's duty is not to echo Cicero and the Ciceronian laments of the pulpit with an 'O tempora! O mores!' but to set about collecting evidence to see what the changes actually were."

[2] "Poverty," he argued, "much as we rightly strive to abolish it, often provides, not a motive for crime, but wholesome incentive. The youngster from a poor home knows he must work hard and conscientiously if he is to gain and keep a job; and this ever-present knowledge at least gives him a purpose in life. Indeed the need to labour, the curse pronounced on Adam has proved one of man's chief blessings." My own experience in following up pupils after their school days are finished has fully convinced me of the truth of these remarks. While I am writing I have in front of me a note from an employer who observes that, "as the risk of unemployment has declined, and as every youth has become assured of a steady and indeed a generous wage, and learns that laziness need not mean dismissal (since his fellow-workers will at once down tools till he is reinstated), his one and only object is, not to accomplish a good job and to uphold the credit of his firm, but to do a minimum of work for a maximum wage. As a result his day-to-day existence has gradually degenerated into a weary search for whatever may take the place of a higher aim—the excitement of betting or gambling, a round of petty sexual adventures, and for the rest sheer boredom. The reduction in poverty, at any rate in this town, has certainly not meant a parallel reduction in vice and crime".

All this of course was in direct contradiction to the inferences drawn from behaviourist theories. If in keeping with behaviouristic theory we deny that moral values have any independent reality, or can be directly recognized as intrinsically valid, then life becomes essentially aimless. Instinct and emotion, the selfish desire for pleasure, for fun, for thrills, and for a good time generally, remain the only objects which the undisciplined child or teenager can be expected to seek. If Mr. Whittaker wants to have this proved by statistics, he will find plenty of confirmation in the book.

I fully agree with Mr. Whittaker that the primary need is ' psychological research on an objective statistical basis '; but, unlike Mr. Whittaker, I think that case-histories must furnish the requisite data' I agree too that the first problem to which such research should address itself is "the nature of so-called values"; and for that reason I feel it odd that no one has taken up a suggestion, thrown out in the very chapter he criticizes, that an illuminating study would be a comparative investigation of a group of exceptionally *virtuous* members of society, side-by-side with a control-group: " the real puzzle ", so the author tells us, " is not so much why some people are wicked, but why so many are good ". As for ' treatment and training ', systematic research is here still more acutely needed; and the most profitable type of treatment for the statistical psychologist to investigate would, I venture to suggest, be not the biological, or the sociological, or the psychiatric methods that are so widely advocated today, but the merits of a specifically educational approach—a study of moral training, which should begin not after the offender has been caught and convicted, but from the very earliest years.

However, as he himself implies, before any investigation is started, we need more preliminary analyses, detailed, systematic, and comprehensive, both of the various concepts involved and the alternative theories available. There have been quite recently one or two experimental inquiries dealing with ' values '. To evade behaviouristic criticism they are usually described as studies, not of judgements of value, but of ' evaluative behaviour '. And, since behaviourism has swept away all interest in traditional theories, the investigators have been left to improvise their own rough and ready procedure. The usual method is just to throw together almost at haphazard a number of ' verbal labels '; the subjects' ' choice-behaviour ' is then arbitrarily assessed; correlations are worked out, and the coefficients are factorized to see whether or not one of the rotated factors may prove to be an ' evaluative factor '. Needless to say, no introspective checks are applied to confirm the proposed interpretations of the factors. Certainly such an experiment is ' objective psychology ' of a sorts; but it can hardly be regarded as scientific research. Let us hope that British investigations, with a more fruitful tradition of statistical research behind them, will plan something a little better than this.

E. D. Williams

Vol. VXI
Part 1
The British Journal of Statistical Psychology
May
1963

THE PSYCHOLOGY OF VALUE

A Reply

By Cyril Burt

Sternel la voce del verace Autore
' Io te farò vedere ogni valore '.
Dante, *Paradiso*, XXVI, 40–2

I. Are Values Objective?

The Multiplicity of Views. Oscar Wilde once announced as the title of his lecture: "Are the Critics of *Hamlet* Really Mad, or Only Pretending?" In a similar spirit one is sometimes tempted to ask, "Are the psychologists who criticize the notion of moral value really as unprincipled as they sound, or are they only pretending?" From the comments of several first-hand observers who have written to this *Journal*—educationists, social workers, magistrates in juvenile courts, as well as from the columns of the daily papers, it is plain that there is a widespread belief that (in the words of one of them, p. 55 above) " the modern followers of Freud and Watson, particularly among the clinical psychologists and psychiatrists who deal with youthful offenders or who seek to guide in public opinion, are largely responsible for the marked decline in moral standards among the rising generation: too often they seem ready to condone what is evil rather than condemn it". Indeed, so severe are some of the strictures, that the behaviourist now runs a risk of being rechristened a misbehaviourist.

Here, however, as mere scientists, we are concerned primarily with the *truth* of the alternative views rather than their possible consequences, though, as we shall see, the two cannot be sharply divorced. In its present phase the controversy had its origin in the sphere not of ethics but of aesthetics. Certain conclusions to which my fellow workers and I were led in the course of various experimental and statistical studies of aesthetic appreciation were subjected to drastic criticisms; the general complaint was that they were " in conflict with the subjectivist or behaviourist trends of present-day theory " (cf. this *Journal*, XIV, pp. 62f. and XV, pp. 163f). But similar objections have since been urged against the working assumptions adopted in our studies of juvenile delinquency and in our supplementary experiments on moral judgements and ethical standards. And so, in Dr. Gregson's phrase, we " come to the underlying issue of *the whole status of value in psychology* ". Much of the discussion, particularly that couched in psychological or pseudo-psychological jargon, is, it must be frankly owned, little more than what Jeremy Bentham used to call ' nonsense on stilts '. Nevertheless, it has apparently had widespread repercussions. Accordingly it would seem high time, as several contributors have recently suggested, that the arguments on both sides should be systematically re-examined, and that whatever empirical evidence is now available should be collated and re-assessed.

Let us start by inquiring in what the " present-day theory ", advocated by our critics, essentially consists. The crucial contention, we are told, is that enunciated by Mr. Paterson at the outset of the discussion, namely, that " as Ayer, Russell, and Ryle have so conclusively shown, the assumption of an ' intuitive capacity ' whereby aesthetic and ethical values are ' directly apprehended ' is an unnecessary and unverifiable postulate which modern psychology has rightly decided to drop ". Dr. Gregson agrees, but holds that Mr. Paterson " fails to press home the case he really has ". All the relevant data, he maintains, can be readily described " without any idea of value as objective "—an idea which in his opinion is " inadmissible in the context of a subjectivist ethics ". The word ' value ' need not be banned. In psychological experiments, he says, we continually find our subjects making statements which ascribe ' value ' to a thing; but such remarks must be regarded " as the way people ' verbalize experience ' ". Words like ' value ' are in effect simply "a terse way of referring to a consensus of choices at a point of time"; the psychologist himself must interpret them "behaviourally". Thus for Dr. Gregson such utterances are just airborne noises; their meaning for the subject himself does not really count.

However, Dr. Gregson's behaviourism is admittedly somewhat diluted. Mr. Whittaker prefers it neat. With the spirited *cocasserie* that enlivens so much of his writing, he sharpens up the argument until it verges on the paradoxical; but at least we know the case we have to answer. "Biologically", he says, "honesty simply *is* the best policy, and nothing more. The invocation of 'value' to lend glamour to our traditional customs and conventions is simply a hang-over from a metaphysical and pre-scientific psychology, harking back to the days when psychology was a mere offshoot of theology—vestigial survivals as antiquated and as superfluous as the caudal muscles that once wagged a human tail . . . Modern behaviourism, profiting by the biological standpoint introduced by Darwin and elaborated by Freud and Watson, has shown that so-called 'ethical' and 'aesthetic' principles are merely the adaptive behaviour of a gregarious and highly evolved primate, which has proved requisite for the survival and progress of his race, and is learnt and relearnt by the conditioning of each individual afresh within the social group in which he is born. The man in the street", he continues, "will no doubt persist in expressing himself in terms of hallowed notions like 'righteousness', 'beauty', and 'truth', with 'happiness' and 'bliss' thrown in; but the psychologist, as a devotee of natural science, must scrupulously eschew all such transcendental concepts as likely to mislead him into countless fallacious conclusions". Thus for Mr. Whittaker, it will be seen, the notion of value is a snare as well as a delusion.[1]

However, all this, as Miss Williams rightly observes, "is not so much an alternative theory as a hotch-potch of theories". Dr. Gregson indeed has advocated at least three different interpretations corresponding to ' three levels of description '. The one point to which all our critics firmly cling is that " beauty, goodness, and the rest are *not objective but subjective*—mere reifications of each person's emotive reactions ". Yet, although this conclusion is put forward as an " empirical result ", it is nowhere supported by factual evidence. The knock-down argument on which they mainly rely is Occam's " principle of minimal assumptions ". " With a little ingenuity all the relevant data ", we are assured, " can be satisfactorily described in terms of the accepted concepts of behaviourism." Hence " the retention of specific ethical and aesthetic notions becomes superfluous, and should consequently be abandoned as involving a ' multiplication of entities beyond what is absolutely necessary ' ".

---

[1] It might perhaps be thought that at the very outset I should provide an explicit definition of what my colleagues and I mean when we use the word 'value': otherwise there is certainly a danger that 'value' may be regarded as a kind of philosopher's stone—an elusive principle, whose sole property is to turn whatever it touches into something precious. Unfortunately, as will appear in the sequel, my difficulty is that value in my view happens to be one of those primitive concepts which are not formally definable. Hence the only way to indicate how the word is used is to invite people to make an inventory of the things they value or prize. So far as I know, the earliest inquiry of this sort, carried out on a rough statistical basis, are the lists obtained by Sir Richard Livingstone from members of his lecture-classes: (*The Greek Genius*, 1912, pp. 126f.). The following are the items which he found most often mentioned, with their relative frequency in brackets: Health (54), Moral excellence or Goodness (47), Affection and Friendship (35), Intellectual excellence or knowledge of the Truth (32), Happiness or contentment (29), Artistic pleasure and Beauty (15), Physical excellence, i.e. bodily appearance, strength or skill (13), Success (13), Work (10), Money (8), Travel (8). A number of somewhat similar lists have since been obtained: (see Burt, *ap.* M. L. Reymert, *Feelings and Emotions*, 1950, pp. 544f. and refs.). The majority of the items so mentioned—e.g., money, health, strength, travel, and most of the additional items which appear in other lists, such as leisure, freedom, long life, a comfortable dwelling—are commonly regarded as *means* to further ends rather than *ends* in themselves, or, to use the customary adjectives, as 'instrumental values' rather than 'intrinsic values'. Others, such as love or affection and happiness, I should regard as reactions to things of intrinsic value. We thus finally come down to the familiar triad of 'essential values' that have dominated discussions of the subject from the time of the later Greek philosophers onward—namely, the Good, the True, the Beautiful, ἀγαθότης, ἀλήθεια, κάλλος (Proclus, *c.* A.D. 440). Among more recent lists probably the one most familiar to psychologists is that of Spranger—"theoretical value or truth, aesthetic value or beauty, social value or love of people, religious value or mystical union, political value or power, and economic value or what is useful" (*Types of Man*, 1928: cf. G. W. Allport, *Personality*, 1938, s.v. 'value' and refs.). Since in the present controversy almost the only values discussed are the ethical and aesthetic, I shall confine myself mainly to these. For the history and philosophical treatment of the subject the reader may be referred to W. Stern, *Wertphilosophie* (1924), J. Laird, *The Idea of Value* (1929), and the writings of W. M. Urban, notably *Valuation: Its Nature and Laws* (1890) and *The Intelligible World: Metaphysics and Value* (1930).

The Limitations of Occam's Razor. Now I am heretic enough to plead unashamedly for what Voltaire has called *le superflu—chose très nécessaire.* It is as necessary in science as it is elsewhere. Certainly, so long as the scientist's aim is merely to *prove* things—that is, to reach verifiable conclusions which can be asserted with the highest degree of confidence —Occam's maxim, I willingly grant, provides a sound methodological canon: the fewer assumptions we include in our initial basis, the more we increase the probability of the inferences deduced from them, and the smaller the target we offer for our opponents to attack. Hence if, as Dr. Gregson and Mr. Whittaker contend, all the psychologist sets out to do is just to " construct operationally testable predictions of human behaviour—forecasts which shall have a maximum degree of accuracy ", then *a priori* concepts and unproved postulates should be reduced to a minimum. But surely the scientist aspires to something more than this; he wants not only to *prove* things, but also to *understand* things. And for this second purpose a far larger array of initial concepts and postulates will nearly always be essential. In other fields, science is, and always has been, full of such speculative postulates—many, if not most of them, little more than refined versions of the familiar notions of common sense. Why then should the psychologist be forbidden the natural wish to explain and to understand as well as to prove and to predict? There is just one needful precaution. Inferences based on the two types of interpretative model—the ' economical model ' and the ' explanatory model ', as they may be called—must always be sharply distinguished, so that the reader should not be misled into attaching to the latter the same amount of assurance that belongs solely to the former.[1]

---

[1] Even here the distinction is not absolute, but rather one of degree. No hard and fast line can be drawn, as Dr. Gregson and others apparently suppose, between predictions which are ' operationally testable ' and ' conjectural explanations ' which are not: the mere fact that an assumption which is not *directly* testable nevertheless serves to increase our understanding of a particular group of phenomena is itself a partial confirmation of that assumption. I may add that the distinction between explanation and proof, for which I have pleaded in the text and in previous publications, is by no means so ' new-fangled ' as Dr. Levi supposes. Every reader of detective stories is familiar with the warning that (as ' Inspector Dawson ' used to say) " while the interpretation adopted by the police provides the only intelligible explanation of the facts collected, it would be hopeless on the present evidence to prove it to the satisfaction of a jury, who will never, or should never, convict unless they are convinced beyond reasonable doubt ". Readers who seek a more formal account may usefully turn to what is, I believe, the most recent exposition of the notion —C. G. Hempel's opening paper on ' Explanation in Science and History ' in *Frontiers of Science and Philosophy* (ed. R. G. Colodny, 1962, pp. 9f. and refs.). " Among the factors that have sustained scientific inquiry", he writes, "are two pervasive concerns: one is man's persistent desire to improve his strategic position in the world by means of *dependable methods of predicting and controlling events*; . . . there is a second—man's deep concern to *explain and understand* the phenomena it presents to him." In a paper, criticizing a recent article of mine on the ' Concept of Consciousness ', Professor Keehn reminds us of Watson's pronouncement on ' the only possible programme for a natural science of psychology ', namely, ' the prediction and control of behaviour ' (' Consciousness and Behaviourism ', (*Brit. J. Psychol.*, LIV; in the press). This formula includes only the first of the two aims mentioned by Dr. Hempel; and Dr. Gregson, we may note, pares down the Watsonian programme still further: in common with earlier psychologists of the " operational school ", he restricts it to " testable predictions based on testable descriptions ".

Practical Implications. The importance of this second and more comprehensive methodology becomes plain beyond all question in the case of those empirical sciences which provide the basis for practical applications. The pure or theoretical branches of science aim at the utmost degree of abstraction and precision, reducing their data so far as possible to a quantitative form, so that their deductions can be made with the accuracy which mathematical or statistical reasoning alone can guarantee: the applied or practical branches are concerned with individuals, whether persons, objects, or events, and their problems are therefore not abstract, but highly concrete. Hence the engineer, the meteorologist, the physician, and the educationist are inevitably obliged to introduce concepts and assumptions which the physicist, the physiologist, or the anthropologist would ruthlessly exclude.

Those who have criticized my " unjustified assumptions regarding the psychology of so-called values " refer chiefly to what Mr. Whittaker calls my " pronouncements as an educational, industrial, and criminal psychologist ". Here, he thinks, I have " continually dragged in concepts and postulates long ago repudiated by objective psychology, which (he says) insists that all such questions can only be scientifically investigated if we treat them as we treat every other problem that arises from the conduct or the predicaments of our fellow human beings, namely, *behaviourally and statistically* ". To this my answer is that, when I am studying one of my fellow human beings, my purpose, unlike that of Dr. Gregson or Professor Keehn, is not merely to " predict or control his behaviour "; it is to understand him. I am not content simply to note his " observable responses "; I want to discover his interests, his range of knowledge, his ideals and aspirations, what are his personal preferences, whether he feels happy, comfortable, or satisfied, and, in the case of younger people, what may be their innate limitations and their latent potentialities. But, like that of the Elephant's Child, my " insatiable curiosity still won't stop ": I want not only to understand other people, but also to understand myself and my own conscious experiences, and (so far as I can) consciousness as such, with all that it implies.

And here, I fancy, we may trace the origin of the battle of wits that has sprung up between Mr. Whittaker and Miss Williams. For Mr. Whittaker, it would seem, as for so many of those who write and lecture on statistical psychology, all the world's a blackboard, and all the men and women merely chalk marks; for him " the end and object of psychological research ", on delinquency or any other problem, is " a set of bivariate frequency diagrams showing the correlation between causal conditions on the one hand and behavioural traits on the other ". For Miss Williams and for those of us who have to deal with actual persons in their everyday surroundings, their quantitative characteristics—those that can be expressed as measurements and plotted as graphs—are by no means the sole objects of our concern. We think of them as living conscious creatures like ourselves, each with his own implicit or explicit aims; and in order to discover their aims, we must first discover what things they value, and what are the priorities in their schemes of value. And whenever the case is one that called for remedial treatment, we too must have our own scheme of values to direct our efforts.[1]

However, in spite of the wide differences that separate us from our critics, there are three matters on which we all are evidently in complete accord: first, that the problems usually grouped together under the traditional labels of ' value ' and ' valuing ' have been sadly neglected by psychologists; secondly, that these problems urgently call for investigation by empirical methods—that is, by experiment, systematic observation, and statistical analysis; thirdly that, before any such investigations are actually planned or carried out, all the available hypotheses should be systematically re-examined, so that the main psychological issues can be clearly defined in advance.

---

[1] Let me say at once that this view is not " merely an abstract philosophical assumption which [we] see fit to adopt ", as Dr. Levi and others seem to suppose. It lends itself readily to experimental and statistical investigation: see, for example, *The Young Delinquent*, 1925, pp. 405 and refs., G. W. Allport and P. E. Vernon, 'A Test for Personal Values ', *J. Abnorm. and Soc. Psychol.*, XXVI, 1931, pp. 233–248, and G. W. Allport and H. Cantril, *ibid.*, XXVIII, 1933, pp. 259–273. It is much to be regretted that experimental studies of this kind have of late fallen out of favour.

Our first task must therefore be to distinguish and classify the various alternatives that have been put forward.

The Classification of Theories. Dr. Gregson's criticisms imply a dichotomous classification of relevant theories into ' objective ' and ' subjective ' respectively. He leaves us rather in the dark about which of the many possible meanings he has in mind for these familiar terms. I myself would prefer to say that the crucial distinction is between (I) those theories of value that regard it as *irreducible*, i.e., as a logically primitive concept which cannot be analysed without remainder into non-valuative characteristics, and (II) those which regard it as *reducible*.

I. My own view is that beauty is an irreducible characteristic which cannot formally be defined, but only intuitively apprehended, like the quality ' yellow ' or the relations of ' before ' and ' after '. And assuming, as in this case I presume we must, that beauty is a *predicable* quality or property (a phrase I use to cover both attributive properties and relational properties), then we may ask whether it is (i) the quality of an *independent object* (not necessarily of a physical object) or (ii) the quality of a *personal experience*. Recently, for instance, in discussing some Venetian mirrors, a colleague told me that he thought " the oval, fifteenth-century specimen far more beautiful than an irregular trapezoidal piece " (of modern workmanship); and, in reply to a further question, assured me that he considered the beauty of the older piece to form " an *objective* characteristic just like its oval shape ". Yet the " beauty " of the second movement of Beethoven's ' Emperor ' Concerto which we heard that evening he held to be an irreducible Gestalt-like quality of the auditory *experience*, and therefore " purely mental, since the actual sounds have no physical existence ": (Dr. Gregson, I take it, would agree with the latter statement, but not with the former.) In either case we may hold either (*a*) that the notion of beauty (and of value generally, including goodness) is an *empirical* concept, like green or loudness, or (*b*) that it is an *a priori* concept, much as Broad, for instance, has argued that ' cause ' and ' substance ' are ' *a priori* concepts '.

II. Theories that maintain that valuative concepts, such as beauty or goodness, can be analysed without remainder into the presence or conjunction of certain non-valuative characteristics are of many different types. The following seem the commonest, though, as we shall see later on, they are not always treated as mutually exclusive. (i) One of the oldest is the *theological* interpretation, invoked chiefly by moralists. Goodness, it has been argued, is simply what is enjoined or sanctioned by God. It is this widespread notion, I suppose, that accounts for the way in which (as Mr. Whittaker observes) so many present day behaviourists—Dr. McLeish, for example—as well as several philosophers in the past, have dismissed the suggestion that morality can be independent of human judgement on the ground that any such assumption is no more than ' an antiquated theological dogma '.[1] (ii) Secondly, there are what may be called *structural* theories, advanced chiefly by art-critics. Some writers, for example, have endeavoured to define ' beauty ' as consisting in certain structural relations exhibited by the object, which together constitute a coherent and harmonious whole, combining a maximum of variety with a maximum of unity. I should myself contend that, though this may be a necessary condition, it is certainly not a sufficient condition, much less that the two are identifiable. (iii) Thirdly, there are *sociological* theories. Many have maintained that what is commonly called 'right' or 'good' is simply a name for the rules or standards which have been enjoined by society or its representatives; and analogous 'cultural hypotheses' have been advanced to account for our ideas of beauty. (iv) Fourthly (and frequently held in conjunction with the previous view), there is the *biological* theory, to which most present-day psychologists lend a ready support. Good conduct is defined as "adaptive behaviour calculated to promote the survival of the individual or his species". Aesthetic qualities are interpreted in terms of sexual selection or the like. (v) Finally, there is a wide variety of *psychological* theories. (*a*) Some hold that the psychological characteristics, into which all forms of value can be analysed are essentially cognitive characteristics, e.g. that "the experience of beauty is essentially a mode of perception in which the spectator's attention is held and stimulated, and yet neither perplexed nor completely baffled, by what he sees or hears". Similarly "moral behaviour" is described as merely "far-sighted rational behaviour". (*b*) Some again hold that value can be equated with certain *affective* reactions, beauty, for example, with what Dr. Gregson calls 'hedonic tone', goodness with what is termed 'true happiness'. (*c*) Others hold that value is essentially reducible to certain *emotional* reactions. According to one version—the 'expressionist' theory—a beautiful picture is one which *expresses* both the artist's emotions and our own; according to another—the 'emotive' theory—it is a picture which we admire or *approve of*, and similarly a good action is an action which wins our approval. (*d*) Last of all, there are theories which seek to reduce value to terms of *conative* tendencies, such as wanting, striving after, or desiring. In the last three cases beauty, it will be seen, becomes a name, not for a simple *quality* or set of simple qualities of which I am directly aware (as I am aware of pleasure, for example), but rather for a *relational property* arising out of the relation between the cognition of the physical characteristics of an object or action and my affective, emotional, or conative reaction or attitude towards it. And in each case there is a cross-classification: the defining reaction or attitude may be just one's own personal experience at the moment, and in that sense *private*, or it may be that of all normal human beings, and in that sense *public:* strictly speaking, as Broad implies, only the former group of theories should be called 'subjective', the latter he would hold are 'objective'.[1]

---

[1] ' Most of us would agree ', said F. P. Ramsey addressing a philosophical society in Cambridge in 1925, ' that the objectivity of good was a thing we had settled and dismissed with the existence of God; Theology and Ethics are two subjects which we realize have no real objects.' A similar view was defended by Ayer in *Language, Logic and Truth* (1936), though he has since explicitly rejected it: (see *Philosophical Essays*, 1954, Chap. X, where he starts by criticizing Ramsey's dictum).

For Mr. McLeish's argument see his recent book, *The Science of Behaviour* (1936), pp. 1f., 16, and 191). Mr. Whittaker declares that " the sheer antiquity " of my view " is sufficient to condemn it". To this we might rejoin that his own hedonic theory is just as antiquated. But the true answer is surely that of the plainspoken critic in Molière's play: " Voyons, monsieur, le temps ne fait rien à l'affaire."

The conclusion to which my own inquiries have led me is that each of the three 'essential values' (p. 60) is an unanalysable quality, directly perceived, but existing independently of the percipient: in short, that

> Value dwells not in particular wills;
> It holds his estimate and dignity
> As well wherein 'tis precious of itself
> As in the prizer.[2]

Every one of the foregoing theories has been touched upon in the course of the present controversy. But for the most part my critics have sought to defend one of the four that I have labelled 'psychological'—sometimes, a little inconsistently, advocating two or more of the others at the same time. Accordingly in what follows I shall confine myself chiefly to these four. None of them is as new as its advocates seem to suppose. Indeed, the older historical formulation of them is usually much clearer and more intelligible than that offered by their present-day supporters. Our next step therefore must be to take each one in turn and define it as unambiguously as possible, and then briefly review the more important objections which have been urged against it. All this will inevitably involve a good many flashbacks into the past.[1]

---

[1] I base this interpretation on his remarks about *ethical* characteristics (*Five Types of Ethical Theories*, 1930, pp. 85, 260). Most psychologists seem to regard *all* 'psychological' theories as 'subjective' theories.

[2] *Troilus and Cressida*, II, ii, 52–5.

[1] The elusive terms 'objective' and 'subjective' have occurred so often in the foregoing discussion, as well as in the previous contributions to the problem, that we should perhaps pause at this point to inquire what precisely these expressions are intended to convey.

Today the word 'objective' has become a veritable shibboleth with diehards of the behaviourist school. Its popularity seems due largely to the writings of Tolman (cf. 'Psychology *versus* Immediate Experience', *Phil. Sci.*, II, 1935, pp. 354f., where it is repeatedly used to describe 'the methods of psychology as an empirical science'). Unfortunately, as a glance at the *Shorter Oxford Dictionary* will quickly reveal, the epithet may be used in at least half a dozen different senses; and none of the psychologists who borrow it deign to tell which of them they accept. Much depends on the meaning implicitly given to the noun 'object'. (i) In pre-Watsonian psychology (e.g. that of Ward), as in epistemology generally, the words 'subject' and 'object' are used as correlatives to designate the referent and the referee of an asymmetrical relation denoted by the term 'consciousness' (or its equivalents); thus "the word 'subject' (it is said) denotes the mind or ego that *has* the experience, and the word 'object' denotes whatever *is* experienced *by* a 'subject'". Whether such an 'object' is an ontological, an epistemological, or merely a psychological object are further issues into which we need not enter here. Since behaviouristic psychologists commonly avoid the word 'consciousness', and reject the whole idea of an experiencing 'mind' or 'subject', this somewhat technical interpretation is no longer open to them. (ii) Nowadays the word 'object' seems to be most frequently used as a synonym for 'material thing', so that an 'objective terminology' means what Carnap and his followers would call the 'thing-language'; and this, we are told, includes not only the names for certain things, but also the names for their observable characteristics—e.g., 'yellow' or 'hot'. On the other hand, we are expressly warned that it excludes such terms as 'temperature', 'weight', 'light-wave'. Now these are quite freely used by the behaviourist in his descriptions of stimuli, whereas he would regard attributes such as 'yellow' or 'hot' as 'secondary qualities' and therefore 'subjective'. Hence most behaviourists would presumably agree with Brunswik that the 'thing-language' as thus defined is "unsuited to supply the language of psychology" (F. Brunswik 'The Conceptual Framework of Psychology', *Int. Encycl. Unif. Sci.*, I, 1952, No. 10). (iii) Among those psychologists who accept the doctrine of physicalism the term 'objective' is often adopted as a convenient single word to designate whatever can be expressed in the language of the physical sciences, which for many physicists would mean whatever can be expressed in numerical terms or reduced to mathematical form. Since Dr. Gregson repeatedly contrasts what is 'subjective' with what is 'physical', and at his highest 'level of description' proposes to express the processes involved in terms of an algebraic equation, it would seem that this is one of the meanings he has in view. Nevertheless, few contemporary behaviourists would accept it as the only type of objectivity, since they want to include in their data such things as verbal communications and reports; and, although these certainly consist of physical movements, it would be quite impossible to express them *solely* in terms of physical movements, and still indicate their nature. (iv) Watson and other behaviourists retain the word 'subject' to denote the person upon whom an experiment or set of observations are made; and the term 'subjective' is then apparently used as a loose description for the processes or phenomena occurring within the space occupied by the subject's body—e.g. such processes as perceiving (or being stimulated), thinking (in the sense of making incipient verbal movements), 'responding emotionally' (which for them implies reactions of the subject's viscera or glands), or showing signs of pain, satisfaction, and the like. The term 'objective' consequently covers any relevant phenomenon occurring outside the subject's skin, and is virtually equivalent to 'external'—a word which is frequently substituted as a synonym. Strictly speaking therefore the colours of objects, as well as the objects themselves, should be classed as 'objective', since they are localized outside the subject's body. When the subject's physical reactions consist of movements or changes that can be observed by the psychologist, they too (somewhat inconsistently) may be termed 'objective', since they are external to the observer. This interpretation is the one adopted by Drever in his *Dictionary of Psychology* ('*objective psychology* is that type of psychology which studies phenomena that can be observed *from the outside*'). (v) Professor Broad would limit the term 'subjective' to those experiences or judgements which are 'private'—i.e., refer solely to the speaker. Thus a subjective theory of goodness is "one which would make ethical judgements consist of statements by the speaker about his own mental attitude at the time of speaking". It

II. SUBJECTIVE AND NON-COGNITIVE THEORIES

1. *Value as based on Desire.* Let us take the last of the foregoing theories first, since that appears to be the view most frequently put forward by our critics, though usually in conjunction with one or other of the rest. "Value", says Mr. Whittaker, "is a quality or name projected on to an object to indicate the fact that it is, temporarily at least, an object of 'desire'; and 'desire' in turn is itself just another name for certain motor responses, which would lead, when finally consummated, to the acquisition of the object, as in hunger or sex. The responses are primarily reflex, but in human beings are subject to widespread conditioning. When overt, they can be observed; when merely incipient, we have to rely on verbal reports."

The earliest formulation of this theory is that of Thomas Hobbes of Malmesbury. "Whatsoever is the object of any man's appetite or desire, that is it, which he for his part calleth *good;* for these words are ever used with relation to the person that useth them, there being nothing simply or absolutely so." Hobbes goes on to argue that desire and aversion are actual "motions", being "either an endeavour toward or fromward something".[1] Later writers extended the theory to the realm of aesthetics: they would say of every form of art what Bacon says of poetry: "it submits the shows of things to the desires of the mind, whereas reason doth buckle and bow the mind to the nature of things".[2]

---

follows that a "psychological theory" of value (e.g. Hume's) need not be a "subjective theory", "for value may still be public" (*Five Types of Ethical Theory*, 1930, p. 260). (vi) With these latter interpretations there is commonly combined the tacit suggestion that what are designated 'objects' or 'objective' form a class of entities or processes which, though necessarily *accessible* to human observation, and in that sense 'public', are nevertheless not *dependent for their existence* on human observation, and furthermore that they are the *only* entities or processes that have these two-fold characteristics. They thus are assumed to possess an 'absolute' reality, which cannot be validly ascribed to any other type of phenomena; all other words and phrases used in this context are mere 'verbalizations', or names which correspond to nothing that is 'real'. This often seems to be the way in which the word 'objective' is employed both by Mr. Whittaker and by Dr. Gregson (who indeed ends by mentioning the 'nominalist-realist controversy'). However, neither they nor the other behaviouristic writers who use the term mention any empirical criterion whereby the presence or absence of this implied 'reality' can be verified. It is therefore a more or less arbitrary metaphysical assumption. Let me add that, however the word may be defined, I myself should be tempted to say that, in view of our limited knowledge, any distinction between what is 'objective' and what is 'subjective' must always be a *matter of degree*, and will vary according to the context: a hard and fast dichotomy of this kind seems out of the question.

[1] *Leviathan* (1651), Pt. I, ch. vi. This comes very close to the interpretation proposed by Dr. Gregson, who maintains that value is not a *quality* of some object, but implies a *relation* between the object and the subject, and then goes on to argue that "we can get round this problem" (the difficulty raised by the positivists when they tell us that any statement predicating value is a meaningless emotive noise) "by saying 'I *like* this shade of red' is behaviourally equivalent to 'this is a *nice* shade of red'". He explains that we can check or improve on the subject's 'verbal reports' about what he likes or desires by observing his overt actions, i.e. what in fact he actually chooses.

Hobbes' doctrine has gained far more support in this country than on the continent. Spinoza, it is true, has occasionally been cited as "an example of a continental philosopher who reduces value to desire or sheer conation". In Part III of his *Ethics* (ix, schol.) he declares "we do not desire or strive after a thing because we judge it to be good; we call it good because we desire it or strive after it". But this view is greatly modified by his discussion of 'the rational good' in Part IV. The supreme good, we then discover, is the intuitive understanding of the whole—the *amor intellectualis Dei*. Thus, as he says in *Letter* xxvii, "ethics rests ultimately on metaphysics".

[2] *Advancement of Learning* (1605), Bk. II, iv, 2.

And so we reach the familiar conclusion:

> There's no such thing as that we beauty call
> It is mere cozenage all;
> 'Tis not the meat, 'tis but the appetite
> Makes the eating a delight.
> If I fancy take
> To black and blue,
> That fancy doth it beauty make.[1]

Schopenhauer in his *Preisschrift* revived Hobbes' doctrine in its naïve form,[2] applying it to both ethical and aesthetic values; in *Die Welt als Wille und Vorstellung*, however, the theory he propounds is rather more subtle (see below pp. 72f.). But among modern adherents probably the most influential has been von Ehrenfels. "We do not" (he writes), "desire or strive after things because we perceive an intangible essence called 'worth' in them; we ascribe 'worth' to them because we desire them . . . Worth or value is a relation between some object and some subject, and expresses the fact either that the subject actually desires the object, or that the presence of the object conditions in the subject an experience higher in the scale of pleasure and pain than its non-existence would do."[3]

To the present-day psychologist such an account is plausible only if, like Dr. Gregson and Mr. Whittaker, he adopts a rigorously behaviouristic standpoint. He hears Tommy say he "wants to win that alley-taw": (I borrow Mr. Whittaker's own illustration); in such a situation there is a 'subject' (Tommy), an 'object' (the marble), and a relation between the two, which the psychologist calls 'valuing', and which Tommy (when he makes what Dr. Gregson calls a 'verbal report') calls 'wanting'. For a strict behaviourist, however, the marble is and can be 'desirable' merely in the sense that it actually is desired. For a psychologist who takes introspection seriously, however, the inferences so drawn are tantamount to arguing that the cart must pull the horse because the tail really wags the dog. "When I myself desire something", a colleague at the Slade had just explained while discussing this very point (we happened to be talking about the purchase of some original etchings by Augustus John), "I desire it because I find it desirable in the sense of '*worthy* to be desired'. I covet it because I consider it to have value; I do not ascribe value to it because I covet it." Each of us, I imagine, could compile a whole catalogue of things whose intrinsic value is unquestionable, which nevertheless he himself does *not* happen to want; and everyone occasionally cherishes desires which do not transmute their objects into things of value; there are evil wishes and malicious intents, as well as yearnings for what is beautiful, good, or true.

2. *Value as based on Pleasure: the Hedonist Theory*. Dr. Gregson, in keeping with many other subjectivists, takes the analysis of the concept of desire a stage further. He does not, like the out-and-out Watsonian, wash his hands entirely of the data furnished by introspection, though he insists, in the behaviouristic portions of his paper, that psychologists must regard such data merely as part of the subject's *verbal behaviour*. He then goes on to argue that, when a 'subject' (who might, I suppose, be a poet or an art-critic) reports that he 'likes', 'admires', or 'wants' such and such a thing, what he is really trying to say is, not that it has any intrinsic merit or value, but merely that "a hedonic tone is pervading the subject's experience of the object"; or, in plain language, that to admire or desire

---

[1] Sir John Suckling, *Fragmenta Aurea* (1646). Some apology is perhaps needed for the way I have interlarded my arguments with quotations; but my purpose is to show that explanations now advanced as new have already had a fairly wide vogue in the past.

[2] "Everything conformable to the striving of an individual Will", he writes, "is called good in relation to that Will." The Will may be that of the group. Thus "an unselfish man is called good by the public because he does what serves the public or conforms with its desires".

[3] *System der Werttheorie* (1897), Vol. I, pp. 2, 65.

a thing is simply to find a 'private joy' in the thought or the perception of it, much as a Bentham's pig found joy in his pig-wash. The flowers of the field, outshining Solomon in all his glory, possess no genuine beauty of their own; it is only our passing pleasure that gilds the lily and lends beauty to the rose.

Hedonism of course is one of the oldest of philosophical theories, both in psychology and in ethics and aesthetics. It was advanced two thousand years ago by the Greek philosophers Aristippus and Epicurus; it was furbished up afresh in the nineteenth century by the utilitarians and their followers. " Nature ", says Bentham, " has placed man under the empire of pleasure and pain . . . His only object is to seek pleasure and avoid pain."[1] To the man of the world this has always seemed a plausible doctrine. To the mathematical psychologist it is equally attractive, because it provides a single universal currency—a hedonic scale, with a negative as well as a positive direction—along which all values can be measured.[2] Nevertheless, the whole theory has been repeatedly subjected to drastic criticism; and today, I should have thought, all who have impartially weighed the difficulties such a notion entails would be agreed that it is quite untenable. To begin with, in its commonest form it involves several flagrant inconsistencies. As ordinarily understood, it embodies two incompatible standpoints. First of all, it is held to imply, as Bentham contended, that pleasure and pain are the sources of all motivation. This the modern behaviourist would in principle accept, though he would prefer to talk of 'positive' and 'negative reinforcement' and leave the words 'pleasure' and 'pain' for the subject's verbal report: in any case he would, like Dr. McLeish, insist that the subject's actions are mechanically and inevitably determined—that he *cannot help* seeking what he calls pleasure and shunning what he calls pain. But secondly, the hedonist goes on to assure us that pleasure—" the pleasure of others as well as of ourselves "—is the only end at which we *ought* to aim. Now if we *cannot help* seeking pleasure, there is no point in telling us we *ought* to seek it. However, let us disregard the first contention, and consider only the second. There are obvious objections. To begin with, whether or not we believe we *ought* to seek it, there is little question that our most intense experiences of pleasure, like our most agonizing pains, commonly overtake us unsought and unforeseen; and as a general rule, it may fairly be said that the best way to *get* pleasure is to *forget* it. Indeed, our happiest moments are usually to be found, not when we attain our end, but during the pursuit of it. " If I held the truth in my hand ", says Emerson, " I would let it go for the joy of seeking it." And few copybook maxims are so familiar as the closing line of Stevenson's *Aes Triplex:* "To travel hopefully is better than to arrive."[3]

According to Mr. Whittaker, however, there is a simple formula. In common with most twentieth century psychologists he believes all motivation can be expressed in terms of a single tendency. " Biologists and behaviourists, psychologists and cyberneticists ", he says, " have combined to teach us that what determines an organism's behaviour can be summed up in one elementary homeostatic law—*the recovery of equilibrium.* The environmental change that forms the 'stimulus' disturbs the organism's existing equilibrium, and its 'reaction' consists in a series of adaptive changes whereby the original condition of stability is restored. This ", he continues, " is all that the behaviourist means by 'need reduction', and all that the psychiatrist implies when he says that 'the basic requirement is security' ". Translated into hedonistic terms, " it means that the only form of pleasure (or rather satisfaction) is escape from pain (i.e. from dissatisfaction) "; translated into valuative terms, it means that a positive value must be defined as consisting merely in the absence of negative values.

This is surely to push the passion for monism and the 'parsimony of hypotheses' to the furthest limit. So long as we confine ourselves to the *experimental* evidence (that furnished by the usual type of laboratory research on rats and dogs) the theory may sound plausible enough;

---

[1] *Principles of Legislation* (1789), ch. I.

[2] For a detailed outline of a 'calculus of pleasure'—'hedonimetry', as he termed it—see F. Y. Edgeworth, *Mathematical Psychics: An Essay on the Application of Mathematics to the Moral Sciences,* (1881).

But there is a further paradox. As expounded by Bentham, the hedonist doctrine enjoins us to seek " the greatest pleasure of the greatest number ". If, however, I am to take into account the pleasure of others, I shall continually discover that this means sacrificing my own. In any case it is easy to think of situations where Bentham's basic principle, taken literally, would unquestionably lead us to choose wrongly. E.g., suppose that (i) I can either (*a*) act so that George gets five units of pleasure, and Tom, Dick and Harry none at all, or (*b*) act so that each of the four receives one unit: maximization, regardless of distribution, would require us to opt for (*a*). Yet this is plainly unjust. (ii) The amount of pleasure gained from women and wine by members of ' Abodes of Love ' is far greater, because far more intense than that gained (with a good deal of pain thrown in) by a team of pathologists investigating the causes of malaria: once again the mere maximization of the quantity of pleasure would lead us to applaud the former rather than the latter.

Mill sought to modify the crudity of the older formulation by recognizing that pleasure may differ in quality as well as in quantity. " Utilitarian writers in general ", he says, " have placed the superiority of mental pleasures over bodily pleasures, simply in their greater permanence, safety, etc., of the former—that is, in their circumstantial advantages rather than their intrinsic nature." He himself holds that " some *kinds* of pleasure are more desirable and valuable: it is better to be a human being dissatisfied than a pig satisfied, Socrates dissatisfied than a fool satisfied; and if the pig or the fool is of a different opinion, it is because they know only their side of the question ".[1] But here the word ' pleasure ' —like so many of the terms popularly proposed as synonyms or substitutes for value—becomes ambiguous. It may mean, and indeed for the psychologist it should mean, what Dr. Gregson calls ' hedonic tone '; in that case it can have no quality but its own. But the word is often used to mean pleasant states, pleasant experiences, or even pleasant objects. And these of course *may* differ widely in quality. But in this case the value depends not on the pleasure but on the quality of the experience or the object.

The objections to the identification of moral value with mere pleasure may perhaps be most succinctly summed up by applying the triple criterion which I have previously

but it is in the sharpest conflict with the evidence from *observation*. We have only to watch the behaviour of our tame pets in the home and wild creatures in captivity to see how false it is. The animals in the zoo, like the domestic cat or dog, live in a welfare state. Well housed, well fed, relieved of the need to hunt for prey or hide from foes, they ought, according to the theory, to be living in a paradise of stable equilibrium. In point of fact, when their environment fails to provide " a stimulus to disturb that equilibrium ", they actively try to create it. For them this comfortable Nirvana means not bliss, but boredom. The same holds good of human beings. As recent research has shown, to deprive an astronaut of all external stimulations is not to reduce him to a condition of comfortable equilibrium, but to set up a series of psychoneurotic disturbances which will compensate for the monotony that ensues. As for children and teenagers, every impartial observer must agree that what most of them seek is not security and ' stable equilibrium,' but adventure.

Mr. Whittaker sums up his theory by saying that " the object of all our efforts is to abolish effort ". I agree that what may be called ' the principle of least effort ' governs many of our compulsory and habitual activities, but it has little influence in activities accompanied by attention, and none at all in spontaneous activities. Even the self-confessed hedonist would admit that the ' pleasures of pursuit ' are more satisfying than the ' pleasures of attainment ', and the former necessarily involve effort. A few have sought to explain the paradox by arguing that the pleasures of pursuit are really pleasures of progressive attainment. But this is by no means true in every instance. We can enjoy a struggle, a fight, or a strenuous game even when we know from the outset we cannot win; and the highest and most satisfying of all pursuits—research in pure science, for example—are from their very nature unending.

I have dwelt at some length on this incidental issue, because the widespread prevalence of the simplified theory that Mr. Whittaker defends is in my view largely responsible for the way psychologists and others, who have been trained along behaviouristic or psychoanalytic lines, fail so frequently in their dealings with delinquent youths and with the moral conflicts of their patients.

---

[1] *Utilitarianism* (1864), pp. 11f.; Mill's italics.

suggested in the case of aesthetic values. First, the variations in the degree of positive pleasure have at best only a low correlation with variations in moral value: some of the most intense pleasures are the most ignoble. Secondly, many things which we all recognize as good are not only unaccompanied by anything that can properly be called pleasure; they may actually involve pain: " Christianity ", said Goethe, " is the worship of sorrow "; and certainly the ' beatitudes ' explicitly preach the value of those states or experiences which are commonly regarded as the reverse of pleasurable. Thirdly, there are many pleasures which are definitely evil, e.g., the *Schadenfreude* of the sadist. Indeed, we condemn malice as a vice, not merely in spite of, but because of the pleasure that permeates it. And finally, does not the essence of the moral problem turn on the conflict between personal pleasure and true morality? " Video meliora proboque," cries Medea, " deteriora sequor ". We see and recognize the good; yet we constantly succumb to temptation.

But today, I should have thought, few psychologists and philosophers would still subscribe to such an identification in the moral realm.[1] Let us therefore turn to examine the hedonic doctrine as applied not to ethics but aesthetics. It was, we may remember, the earlier discussion of ' aesthetic factors ' (this *Journal*, XIII, pp. 87–92; XIV, pp. 72–77) that gave rise to Dr. Gregson's criticisms of my views on value generally. His account receives some support from the dictionary: " Beauty ", it tells us, " is that quality or assemblage of qualities which gives the eye or other senses intense pleasure."[1] Once again the theory goes back to the Greek philosophers. Plato in the *Hippias Major* says that many of his contemporaries described beauty as a kind of pleasure, and that Socrates defined it as that which gave 'profitable pleasure'. Plato himself entertained a very different view. For him the two forms of value—ethical and aesthetic—were intimately allied.[2] He would have his citizens surrounded from their earliest years with objects of beauty, so that they might eventually be led to prize the loftier beauty of holiness, goodness, and reason.

---

[1] Some comments on the foregoing arguments, sent me by Mr. Paterson, call for a brief reply. He " willingly endorses [my] contention that behaviourism and hedonism are mutually inconsistent ", but maintains that " for one who like [himself] repudiates a purely behaviouristic psychology, [my] criticisms lose their force ". " I should agree with Mr. Whittaker ", he writes, " that ' the kind of conduct which Tommy's mother calls good is the kind of conduct that pleases her ', and further that, if we ask why *that* particular type of conduct pleases her, the answer is, she believes that it will in the long run make for Tommy's happiness. But ", he continues, " though pleasure and happiness are motives of behaviour, they are not *modes* of behaviour; they are *feelings*, and as such wholly subjective. Hence goodness and wickedness (and for much the same reason beauty and ugliness) are likewise subjective. Thus, I accept the conclusions which Dr. Gregson and Mr. Whittaker have drawn, though I reject their arguments."

To illustrate his point, he appends one or two examples. " Tommy sniffs. By itself the action is indifferent. But the constant sniffing displeases his mother. It therefore becomes wrong. The same holds of hitting or stealing, except that for ' mother ' we must substitute ' victim '. In itself ' the primrose by a river's brim ' is ' a simple primrose . . . nothing more '. But it gives pleasure to the poet, and therefore becomes beautiful. Thus, action *plus* pleasure = good behaviour; object *plus* pleasure = beautiful object."

Evidently then for Mr. Paterson there is nothing either good or bad but feeling makes it so. This view, however, plainly leads to much the same paradoxes as the version of hedonism which Mr. Whittaker supports. We cannot possibly suppose that, when Loyola " renounced the pleasures of nobility and wealth ", or St. Perpetua " chose death rather than the delights of a pagan life ", their motives were " some kind of pleasurable feeling ". Nor can the well-known maxim " It is better to suffer wrong than to do wrong "—a principle preached and practised by Socrates as well as by Jesus—be " reduced to hedonistic terms ". Where then do the fallacies lie? The first step in Mr. Paterson's argument, like that of most hedonists, involves the stock fallacy of " inverting the causal direction ". Tommy's behaviour pleases his mother because she considers it good; she does not consider it good because it gives her pleasure. In the second step the issue as to whether or not Tommy's happiness is, from Tommy's point of view, just " a feeling, and as such wholly subjective ", is quite irrelevant. For Tommy's *mother* her son's happiness is evidently an *objective* possibility.

Similarly in the field of aesthetics: what pleases me in a ' hedonic situation ' (to borrow Mr. Paterson's phrase) is nearly always some *non-hedonic* characteristic. The beauty of the primrose is not just the temporary effect of *adding* my pleasure (or the poet's pleasure) to the flower: the beauty is, so to speak, a ' public ' quality, not a mere ' private ' feeling. To argue along the lines laid down by Mr. Paterson is (as Eric Gill once said of a similar theory) " to treat beauty as a kind of sauce added to an otherwise stale and unpalatable dish. A good dish needs no sauce; a good picture is pleasing because it is a good picture ".

In modern times the strongest support for a hedonic theory of beauty has come mainly from psychologists. "This", so Dr. Levi assures us, "is the theory which the whole psychology of aesthetics from Fechner to Valentine seems to confirm; beauty is just a popular name for the pleasurable hedonic tone which suffuses certain intellectual experiences; and those individuals who rank high in the measurement for the 'general factor of aesthetic capacity' do so because by their preferences they indicate that they are the most sensitive to this type of pleasure." It is a view that was shared by a small coterie of poets, artists, and art critics towards the close of the nineteenth century;[3] but it is entirely out of keeping, not only with the attitude which most writers on aesthetics—whether subjectivists, objectivists, or expressionists—have adopted during the present century, but also with that which has dominated the general trend of art criticism throughout the past. This alternative view, with its frequent comparison of beauty and goodness, comes much nearer to the doctrine first formulated and developed by Plato and his Neo-platonist followers. We have only to recall the testimony of Dante in the *Commedia* and the *Convivio*, Spenser in his *Hymne of Heavenlie Beautie*, Shelley in his *Defence of Poetry*, Wordsworth in *The Prelude*, Ruskin in *Modern Painters*, and Bridges in his *Testament of Beauty*.

However, instead of warming up stale arguments, let us glance at one or two concrete examples. When we watch a tragedy like *Hamlet* or *King Lear*, the feeling which is uppermost is hardly to be described as pleasure. It is excitement, tension, often indeed (as Aristotle held) sheer terror and pity. The agony aroused may become so acute that children often take a lasting dislike to all forms of tragedy—as did one of my own young subjects, who remarked that she "had so hoped it would all wind up with a happy ending; why couldn't Hamlet have married Ophelia, and then forgiven the king and queen?". Or consider a typical painting from the brush of Picasso, say the *Woman in an Armchair*. Even Dr. Gregson would hardly try to persuade us that its salient characteristic is the "pleasurable hedonic tone" with which the viewer's experience is suffused. The savage displacements of the limbs, the violent distortions of their shapes, the arbitrary conjunction of incompatible points of view—all seem deliberately designed as a vehement protest against the idea that the artist's overriding function is to give mere pleasure. The whole posture suggests a body writhing in anguish. And the intention is further underlined by the fact that the twisting of the arm over the head is the classical symbol of pain. Yet surely no serious art critic could deny the picture's claim to genuine 'aesthetic value'.

---

[1] Cassell's *English Dictionary*, s.v.; *The Oxford English Dictionary* uses the word 'delight'. Among modern philosophers Santayana is perhaps the most eminent supporter of a hedonic theory. "Beauty", he says, "is pleasure objectified—pleasure regarded as the quality of a thing . . . Apart from ourselves there can be, in a mechanical world, no element of value whatever: in removing consciousness we remove the possibility of worth" (*The Sense of Beauty*, 1896, sect. ii, 'The Definition of Beauty').

[2] The Greek word καλοκἀγαθία expresses the affinity between the two. Cf. *Republic*, III, 401 A–D. Rightly understood, the attack in Book X on painting and poetry as imitative is fully in keeping with the standpoint I am here maintaining.

[3] I am thinking of the 'ultra-poetical super-aesthetical' l'art-pour-l'artism of Gautier and his followers, culminating in the manifesto appended by Pater to his *Studies of the Renaissance*, (1878). It became part of the 'cult for culture' burlesqued by Mallock in *The New Republic* and by Gilbert and Sullivan in *Patience*. But it is only fair to remember Pater's protest: "I wish they would not call me a hedonist; it produces such a bad effect on people who know no Greek."

I have found it an interesting experiment to submit to my various subjects the same painter's well-known series, *Les Deux Femmes Nues;* the first portrait—a delicate satire on the traditional conception of the nude as a delicious human shape—is the almost invariable choice among those who equate the beautiful with the pleasing; the last two versions are those selected by nearly every person, whether connoisseur or otherwise, who ranks high in our 'measurement for the general aesthetic factor'. The same holds good in other fields of contemporary art. "One of the distinctive features of modern painting, modern sculpture, modern poetry, and modern music", says James Wilson, "is its manifest reaction against the sentimentality and the saccharine sweetness that the romantic school had mistaken for art."

I recognize that some degree of pleasure (or, as I should prefer to say, satisfaction) usually accompanies the experience of objects that have worth or value. But the pleasure cannot serve as an index or criterion of the value, much less be identified with it. In the first place, there are many pleasant experiences—a hot bath, a cold shower, the smell of roasting coffee beans, the taste of mutton and mint sauce—that no one would call experiences of beauty. Perhaps then we should limit it to pleasures from the higher senses or from intellectual activities. Beauty, says Thomas Aquinas, is "id quod ipsa apprehensio placet" —that which affords pleasure through sheer disinterested contemplation. But again the definition covers far too wide a range. I was recently invited to inspect a model dairy farm: its manifest efficiency gave rise to a genuine pleasure arising from "sheer disinterested contemplation". Yet I doubt if anyone would have classed it as an object of beauty. Nature seems to have annexed the strongest pleasure, not to those things whose value is commonly described as 'intrinsic', but rather to those that have an instrumental value —particularly the personal and egoistic activities that are essential for preserving either the self or the race.[1] To suppose that the value of something good or beautiful consists solely in "the pleasure that we subjectively feel in contemplating it" is rather like supposing that the truth of an elegant mathematical demonstration consists solely in the pleasure we experience as we follow the ingenious reasoning.

3. *Value as Expressiveness.* Among those who are themselves creators of works of art—particularly the painters, musicians, and poets of the romantic and neo-romantic schools—one of the most popular notions is that beauty is a mode of self-expression. Of all the theories here discussed it is the one that had exercised the greatest influence on the development of modern art. Its supreme philosophical exponent is Schopenhauer (*Die Welt als Wille und Vorstellung*, 1819, III); but a more up-to-date version of it has since been elaborated by Benedetto Croce.

According to Croce art is "a language expressive of feeling". "A picture, a poem, or a sonata is beautiful to the man whose feelings are expressed in it." When we look at the picture or listen to the poem or the sonata, we recognize in it emotions which we attribute to a fellow-creature like ourselves who expressed what he felt by painting the picture, or by writing the poem or the sonata. The " true function " of a work is thus " to express emotion ". But there is no need for it to represent natural objects in order to have something that will serve as intermediary vehicles for the emotion: the emotion can, so to speak, be delivered direct from the manufacturer to the customer. Nevertheless, though what we really contemplate is volitional or passionate, the aesthetic experience as such is ' theoretical ', i.e., cognitive rather than emotive: the *esse* of beauty is *cognosci*. Since the artist's own experience is always individual, it is incapable of description in scientific terms; hence we recognize it, not by means of our reason or our understanding, but by a kind of empathy, or, as Croce prefers to say, ' intuition '.[1]

---

[1] Herbert Spencer, and the later psychologists of the utilitarian school who adopted his evolutionary theories, maintained that *all* values are instrumental values or the result of associations with such values. In origin they owe what we now call their value to the way they have facilitated, and often still facilitate, the adjustment of the individual and the group to the environment—particularly (in the later stages) to the social environment, and so make not only for the survival but also for the evolution of the species. With later writers this biological interpretation was frequently combined with Schiller's well known 'play-theory' and with Freud's so-called 'pleasure principle'. In this way, it was held, both ethical and aesthetic judgements could readily be derived from non-ethical and non-aesthetic principles. As an account of the origin of such judgements the theory no doubt embodies an important element of truth. But the inferences commonly drawn rest on one of the commonest fallacies, namely, the assumption that, in describing the *origin* of a thing, we have also explained, or explained away, its *validity*. Because chemistry has developed out of alchemy and astronomy out of astrology it does not follow that the validity of the speculations of the modern chemist and the modern astronomer can therefore be rejected without further examination. Today such evolutionary interpretations still seem popular with anthropologists and sociologists, but are relatively rare in psychological writings. They have not been advanced by any of my own critics, and so call for no detailed discussion here.

Among psychologists, save for the observations and inferences of certain leading psychoanalysts, this view has been unduly neglected.[2] As expounded by psychoanalytical writers the expressionist theory has a double application ; it maintains, first, that " in creating a work of art—whether picture, poem, or musical composition, the *artist* is primarily expressing his own emotions, often quite unconsciously ", and secondly, that " in looking at the painting, reading the poem or hearing the music, the *spectator or listener* interprets it, again to a large extent unconsciously, as an expression of *his* emotions ".

(i) The first part of the theory is commonly supported by a variety of clinical observations. Time after time, we are told, the psychoanalyst finds that what a patient writes or draws or paints, both by its content and by its style, throws a flood of light on his conscious and unconscious interests and motives; and this, it is argued, implies that " his interests and motives, and indeed his whole history and social background *find a condensed expression* in his verbal or pictorial productions ". Nevertheless, such observations scarcely justify us in erecting this useful psychodiagnostic principle into a theory of normal art-production. After all in this broad sense almost everything we do or say is a form of self expression. Scientists, like Galileo and Newton, no doubt are partly influenced in their choice of topics and treatment by the social conditions of their time and the circumstances of their own past life and individual temperament; but in the main their choice expresses " their current feelings, interests and problems ": nevertheless that does not warrant our declaring that " the function of a work of science is to express emotion ". In spite of Dr. Rosenthal's illuminating ' depth-analysis ' of Turner's Alpine paintings, an untravelled analyst would, I am convinced, learn far more from them about the Alps and their beauty than about Turner and his complexes. The emotional characteristics of the artist (other than his aesthetic experiences) are but one incidental factor in the work that he produces.[3]

---

[1] B. Croce, *Estetica* (1902), esp. chaps. XIII–XV. Since the first function of language (as the modern theory of linguistics assures us) is communication, i.e. the transmission of experience, it would seem that in contemporary terms Croce's theory is really a theory of art as ' communication ', of which ' expression ' is merely a limiting case. When a man gives vent to his feelings by means of words, whether in prose or in verse, he nearly always assumes that someone besides himself will hear or read what he has uttered; and few paintings are pictorial soliloquies.

[2] The psychoanalytic theories started with Freud's famous study of a painting by Leonardo, Ernest Jones' study of *Hamlet*, and Karl Jaspers' study of Van Gogh. In what follows I am much indebted to Ernst Kris, *Psychoanalytic Explorations of Artistic Vision and Hearing* (1953), and above all to E. H. Gombrich, *Art and Illusion* (1960), although I myself cannot agree with their own conclusions.

[3] According to the psychoanalyst the painter ' interprets ' the landscape before him much as a patient interprets a Rorschach inkblot, and in both cases, it is argued, the interpretations " tell us far more about the interpreter than they do about the object interpreted ". The device of using blots and splashes to stimulate the imagination is at least as old as Leonardo. In the 18th century Alexander Cozens wrote a manual on *A New Method of Assisting the Invention in Drawing Original Compositions in Landscape* (1785); but the landscapes which the author himself claims to have seen in the blots he reproduces are remarkably like the type of landscape then made fashionable by Claude

(ii) What the *spectator* sees in a picture, we are told, is obviously influenced by his own emotional attitudes and predispositions. But, as Professor Gombrich has so clearly shown, the spectator's interpretation depends much more on the mental stereotypes—the 'apperceptive schemata', I am tempted to call them—which he automatically applies when he recognizes what the picture is intended to represent; and these schemata tacitly presuppose the techniques and the conventions regularly used in paintings with which he is already familiar.[1] To those accustomed to the style of Cimabue and his school the 'modern manner' of Giotto (so Vasari tells us) seemed startlingly natural.

Lorraine. On the other hand, in the following century Justinus Kerner (the physician, poet, and spiritualist, who today is chiefly remembered for the famous student song "Wohlauf noch getrunken . . . Juvivallera") used inkblots (*Klecksen*) on folded paper as a stimulus for his poems (*Klekso-graphien*, 1857). Being himself an ardent spiritualist and an intimate acquaintance of Frederika Hauffer, 'Die Seherin von Prevorst', what he usually saw in the blots were demons and the spectres of departed friends.

Turner used to give a paint-box to small children and watch what novel shapes or colour harmonies would appear in their random daubs. Klee and Picasso formed the habit of "doodling to see what would turn up". Galton relates that, having heard of Turner's experiments, he himself deliberately practised "the automatic construction of figures by allowing my hand to scribble at will while I gave my best attention to what was being said by others"; he notes that "it was always a surprise" to see the fantastic results that eventually emerged.

In what follows my intention is not to decry the method as a useful instrument for exploring individual differences. Indeed I myself long ago used ink-blots and pictures, ambiguous and otherwise, as a method of probing emotional tendencies and complexes: (cf. *Mental Differences between the Sexes*, J. Exp. Pedag. I, 1912, pp. 273f. 355f.; also Amya Sen, *A Statistical Study of the Rorschach Test*, this *Journal*, III, 1950, pp. 21f. and refs., and *id.*, *A Preliminary Study of the Thematic Apperception Test*, *ibid.*, VI, 1953, pp. 91f.). The notion that gazing at a set of pictures may precipitate a revealing 'projection' of the spectator's deeper emotions is by no means a new discovery. In Shakespeare's *Rape of Lucrece* the victim's eyes light on "a piece of skilful painting", depicting scenes from the siege of Troy; "much imaginary work was there . . . A hand, a foot, a head stood for the whole to be imagined . . . And so Lucrece, set a-work, sad tales doth tell to pencill'd pensiveness and colour's sorrow. She lends them words, and she their looks doth borrow". And who can forget Polonius's apperceptions of the cloud, as a camel, a weasel, or a whale, or Hamlet's confrontation of his mother? "Look here, upon this picture, and on this."

[1] As we have learnt from studying the development of drawing ability in children, the youthful draughtsman's first efforts begin with what are in effect descriptive schemata or symbols representing, not what he sees, but his conceptions of what he sees. And when he strives after greater realism, he learns far more by copying the drawings in his illustrated story-books than by trying to reproduce the objects he actually beholds. When later on he attempts to sketch from nature, it is still the schemata that fit his own apperceptive images and his own habitual techniques that stand out and catch his attention: (see Burt, *Mental and Scholastic Tests*, pp. 417–429 and figs. 53–64). And the same holds good (as Nietzsche ironically remarks) of the professional artist.

Treu der Natur und ganz! Wie fängt er's an ?
Er malt zuletzt davon, was ihm gefällt.
Und was gefällt ihm ? Was er malen kann!

The catalogues of the larger art libraries include the titles of innumerable drawing books and manuals of painting going back to the 16th century and earlier. Even the most original artist has always begun by appropriating from his predecessors a kind of mental pattern book containing useful pictorial schemes, and then refined upon them. These he uses to classify, according to *genus* and *species*, what he observes; he then modifies, corrects, and adjusts the scheme to represent the *individuality* of the objects and the scene before him. Once more the basic tendency is neatly described by a poet, W. H. Auden:

An abstract model of events
Derived from dead experiments
Becomes, though still particular,
An algebraic formula.

Credette Cimabue nella pintura
Tener lo campo; ed ora ha Giotto il grido,

says Dante. We ourselves, with Raphael's Madonnas in our minds, find Giotto's 'manner' archaic and stiff. If Gainsborough's pictogram for a tree corresponds with ours, we think it marvellously realistic; and by contrast Constable's 'snow' appears a very unconvincing expedient for suggesting the glitter of raindrops on the leaves of British elms. Then, we gradually learn to accept and to admire Constable's conventions. Later still, as we turn from Constable to the landscapes of Corot, Monet, Cézanne, Utrillo, or Hokusai, we have in each case to increase our apperceptive resources and learn a new pictorial vocabulary and grammar. Most of us who in our youth were inured to the 'kanons' or norms laid down by Pheidias and Praxiteles have needed a good deal of rehabituation before we could appreciate the genuine beauty of the best Egyptian sculpture or the statuary of Henry Moore. Thus, if art is 'expressive', what it expresses are the artist's schemes or conceptions quite as much as his emotions or feelings; and yet, at its best it can, by means of these conceptions and emotions, effect a unique poetic evocation of what is (I maintain) objectively beautiful.

The experienced artist himself is well aware of the extent to which he can rely on the interpretations of the contemporary spectator. Leonardo, the pioneer of naturalistic painting, knew how much a picture could gain in realism when some of the realistic details were left out or reduced to the barest hints. Even the most casual visitor to our art galleries cannot fail to note how, in Hals' *Laughing Cavalier* or Sargent's *Carmencita*, what seem, while he stands near enough to 'smell the paint', to be simply slapdash brush-strokes, miraculously resolve themselves at a few yards' distance into a rich embroidery or an intricate lace. And today, the *sfumato*, the *bravura*, or the *sprezzatura*, which were once the hall marks of skilful painting, have turned into mere technical tricks which every twentieth-century caricaturist and advertisement designer will exploit *ad nauseam*.[1]

I should make much the same reply to the expressionist theories of music and architecture developed still more recently both by writers on aesthetics and by practitioners themselves. But a detailed discussion would prove rather more technical and require more space than is practicable here.[1]

---

[1] The way in which what we mistakenly suppose to be direct perception proceeds by filling out fragmentary clues is admirably exemplified by our experiences at the theatre. Seen from the back of the stalls, the heroine looks exquisitely beautiful; inspect her through your opera-glasses, and too often the illusion is completely shattered.

The ambiguities to which these devices may give rise when pushed to their furthest limits, have of late been ingeniously developed by the Cubists, the German Expressionists, and the various other schools into which the modernist movement has eventually split—most of all perhaps by the socalled 'psychoanalytical art' conceived by that *enfant terrible* of Surrealism, Salvador Dali. He begins, like Picasso, by breaking up the basic unit of our daily practical world—the distinguishable 'object' or 'thing', thus splitting the humdrum atom, as it were—and leaves his explosive extravaganza without any single eye-catching element to focus the attention. All natural shapes are reduced to their simplest geometrical forms, and squashed flat against the picture-plane and the various disconnected fragments are thrown together with all the inconsequence of a dream. The spectator—when he gets over the first visual shock—is thus left to tease himself by apperceiving or interpreting what he sees first in one way and then in another; and so in the end may (or may not) find that, in some subtle or sophisticated fashion, they express certain inexpressible or unsuspected feelings of his own. As he stares at these 'cryptograms on canvas', the critic who has been brought up on more orthodox styles may well be tempted to echo the Devil in Kipling's *Conundrum*:

The Devil whoops, as he whoop'd of old:
"It's clever! But is it Art?"

I fully agree with Dr. Levi in deploring the way psychologists have neglected 'the rich field of modern art and music.' The problems are too complex to discuss in detail here. I can only hope that the brief remarks that follow will suffice to dispose of the argument that "subjectivity of modern art is enough to show that artistic values are essentially subjective".

The fatal objection to a purely expressionist theory of beauty lies in its inapplicability to the contemplation of Nature. As an explanation of our reactions to the work of another human being, blessed with feelings and emotions similar to our own, the theory may seem plausible enough; but it cannot account for our discernment of beauty in a flower, a landscape, or a night of stars. Partly no doubt for this reason Croce, like Kant and Hegel before him, deliberately excludes Nature and natural objects. To compare the Venus of Milo with a landscape by calling both beautiful, he says, would be " blasphemy ". Yet could she not be compared with a beautiful woman? One or two writers, it is true, have maintained that " we only find Nature beautiful in so far as she reflects our own emotions ". But such an argument is tantamount to erecting what Ruskin[1] has taught us to call ' the pathetic fallacy ' into a universal principle. Moreover, it is an inversion of the expressionist theory in its original form. According to that the beholder sympathizes with the feelings latent in what he beholds; and now we are required to assume that what he beholds in Nature appears to sympathize with the feelings latent in the beholder.

---

[1] Roughly synchronizing with the various post-impressionist movements in painting, i.e. just before the first world war, there was a determined attempt to win a greater freedom in music by breaking up the units of the traditional style and abandoning the familiar organization into readily recognizable keys, rhythmic patterns, melodic outlines, and harmonic sequences. Bars and key-signatures were deleted; the rules of consonance and harmonic progression defiantly violated; and a 12-note scale introduced to obviate tonality. The ideals of the romantic school a century before were thus carried to the furthest limit: the aim of music, as of painting, was to express the 'inner experiences' of the artist. Scriabin, for example, in his later works (e.g., the *Poem of Ecstasy*) evolved a 'mystical chord' of six superposed 4ths (four perfect, two augmented), and even projected a 'Mystery', in which dance, speech, music, colour, and perfume were to be combined; the results were to embody the 'theosophical illumination' derived apparently quite as much from the writings of Madame Blavatsky as from first hand 'ecstasy'. The leader of the 'expressionist' school was Schönberg—a friend of Kadinsky, the founder of expressionism in painting and later (with Klee) professor at the Bauhaus. Schönberg was himself a painter in the same style.

The first impression conveyed by the new music was one of complete individualistic anarchy. But closer study showed that the rupture with tradition was far from complete. Scriabin himself complained that he could never escape the implications of the dominant 7th. Schönberg supplied a textbook, beginning with a masterly analysis of classical harmony, and then attempting a theoretical basis for the new style. Rigorous rules were prescribed, e.g., that no note can be repeated until all the others in the chromatic scale have been sounded. And, when all else failed, 'the unconscious' was freely invoked to lend an artistic verisimilitude to an otherwise bald and unconvincing explanation.

Thus, here as elsewhere, *order insists on reappearing*. With patience and concentrated study the sympathetic listener gradually feels his way to the schemata implicit in each composer's works (often traceable to earlier sources, notably Liszt, Wagner, and Debussy); and with these additions to his apperceptive resources, he becomes able to appreciate the almost neurotic sense of tension, conflict, and revolt, which so much of the earlier music expresses, and to enjoy the novel possibilities opened up by these daring experiments, ready for later composers in the modern style.

Somewhat similar developments took place in architecture. But, partly as a result of the general impoverishment that followed the first world war, especially in Germany, the *emotional* expressionism of the pre-war period gave way to a *functional* expressionism, in which the aim was to express, not the artist's or the community's conscious or unconscious emotions, but rather the most economical adaptation of the work of art to its specific purpose. The highest type of value was a utilitarian or instrumental value rather than a psychological. Artistic creation itself was now subjected to Occam's razor. Art and technology—particularly structural engineering—became allies rather than enemies. And here to a large extent the severely abstract tendencies of Bauhaus architecture ran clearly parallel with the desire for 'abstraction' instead of 'representation' in painting and music.

Today, however, in each of the fine arts we may, I fancy, detect a reaction and an incipient return to traditional principles in a more or less modernized or sophisticated guise. Klee in his latest works, as his posthumous *Bildnerisches Denken* (1956) explains, came to regard his art as "an attempt to communicate a world as it might be or is in process of becoming". Klee and Kandinsky were musicians, and gave *musical* titles to their works, for music is abstract. But in the past the most influential painters—Michael Angelo, Blake, Turner, Rosetti, and the best known Chinese artists—were more often pre-eminent as *poets*. So, too, in a minor way are several of our contemporary innovators. Indeed the prolonged taboo on 'literary associations' is at last being lifted once again; and, while I am writing this note, there is in London an exhibition of work by Kurt Schwitters which he entitles 'The Visual Images of a Poet',—hailed by the critics as "typical of the new revival—a gradual reappearance of the concrete in place of what is purely abstract". Thus the order emphasized by the abstract painters is revealed once again in the concrete.

Bosanquet, however—the only philosopher of eminence who has attempted such an extension—has given the theory a slightly different turn. He proposes to define beauty, both in art and nature, as " that which has individual expressiveness ".[2] But from the explanation that ensues it seems that by *expressing* our emotions he really means *stimulating* our emotions. And then we immediately inquire why some objects or scenes should stimulate our emotions in such a way as to excite aesthetic appreciation while others do not? The answer presumably is contained in the phrase " conditions for general or abstract expressiveness "; and when we ask what precisely this includes, we find after all that it covers almost exactly those qualities or conditions which believers in objective beauty ascribe to beautiful objects, and that these are ' directly apprehended '.

4. *The Emotive Theory of Value.* There is, however, yet another type of ' affective subjectivism ' which has recently come back into favour. Hume in his *Treatise of Human Nature* (Book II ' Of Morals ', 1739) endeavoured to show that " moral distinctions cannot be derived from reason ". If, for example, we argue: " Working at a useful occupation is a method of increasing useful commodities; increasing useful commodities is a result which will bring pleasure to a large number of people; therefore I ought to work at a useful occupation ", then (he maintains) " this *ought* expresses a new relation or affirmation which cannot be deduced from propositions which merely contain *is*." Hume therefore concludes that assertions which contain *ought* or *ought not* do not state reasoned inferences from facts—though reasoning may be needed for a preliminary analysis of the facts; they merely express " emotions of approbation or disapprobation ". And he adopts much the same theory to explain statements about what is or is not beautiful. But at this point we should perhaps recall Wordsworth's well known comment—that " Hume was the worst critic that Scotland produced ".[3]

Hume's theory has since been revived and popularized by Professor Ayer. " Aesthetic words, such as ' beautiful ' or ' hideous ' ", he writes (*Language, Truth and Logic*, 1936, pp. 109f.) " are employed, as ethical words are employed, not to make statements of fact, but to express certain feelings and evoke a certain response." And once again we ask, what can these feelings and responses be if they are not just feelings or responses of pleasure? In the case of ethics, we are told, there is "a special sort of *moral* approval or disapproval which is the feeling that is being expressed". In the case of aesthetics it is presumably a special sort of *aesthetic* 'approval or disapproval'.

---

[1] *Modern Painters*, III, iv, chap. xii (The Pathetic Fallacy).

[2] ' On the Nature of Aesthetic Emotion ', *Mind*, III, no. 10.

[3] What Hume is really attacking are theories like those of Clarke and Locke (Essay, III, xi, IV, iii), who had sought to refute Hobbes by showing that moral duties were capable of rational demonstration. He himself is to some extent influenced by both Butler and Hutcheson (see below pp. 83f.), and is in effect attempting to reconcile the notion of a ' moral sense ' (see title to Sect. ii of Bk. III) with a utilitarian theory not unlike that of Bentham and Mill. His final views are most clearly expressed in his *Enquiry concerning the Principles of Morals* (1751), which he described as " of all my writings incomparably the best ". Aesthetic values, he tells us, are on much the same footing as moral values. In the *Treatise* (Bk. II, Pt. 1, sect. viii), it is true, he at first says " beauty, like wit, cannot be defined, but is discerned only by taste ". But elsewhere he argues that it is a kind of implicit utility, and so produces pleasure in the beholder. " Most works of art are esteem'd beautiful in proportion to their fitness for the use of man." In his essay *Of the Standard of Tast* he says of aesthetic appreciation, as he has done of moral judgement, that " the principles of taste are universal, and nearly if not entirely the same in all men, though few are qualified to give judgement on any work of art ".

Now as a psychologist I find it difficult to regard approval or disapproval as a feeling or an emotion. To approve, so the *Oxford Dictionary* tells us, is "to pronounce or consider good"; and to do that is surely to make or utter a judgement, not to "express a feeling". And this is the view of so careful a psychologist as McDougall. "Approval and disapproval", he observes, "have been treated by Westermark and others as emotions. But to describe them as emotions is to perpetuate the chaos of psychological terminology. They are not emotions but judgements." Like other judgements, they may be partly determined by emotion; but they may also be unemotional and purely intellectual conclusions, "made in logical accordance with previously adopted principles".[1]

However, let us waive these points of psychological nomenclature, and consider what precisely the theory implies. And at the outset we naturally inquire—whose approval or disapproval? In the case of ethics there is a pretty wide choice: it may be the approval (i) of the speaker himself (Ayer, for example, commonly talks in the first person singular), (ii) of 'us', i.e., the particular philosophical group which the speaker represents (usually designated without clear specification by the first person plural), (iii) of all mankind, except perhaps for a few abnormal deviants, such as defectives, lunatics, and criminals, (iv) of all persons of sound moral judgement, (v) of the majority of those in our own social group, (vi) of those who in virtue of their class or office can authoritatively decide what conduct should receive approval—the priests, the prophets, and the parsons, the legislators and the judges, the powerful and the privileged, or possibly those who hand on to each succeeding generation the accumulated results of past traditions—our elders and our parents, and finally (vii) of God Himself. Roughly speaking, the interpretations might be grouped together as the aristocratic, the democratic, the oligarchic, and the monarchic or theological. Each of these alternatives has had its advocate in the past; but there is no need to examine all of them in detail here. None of my critics takes his stand on the last two alternatives. Several, however, appear to favour what may be called the democratic alternative; and, since no one today would hold that all mankind except the criminals, lunatics, and defectives are agreed in what they approve, we should perhaps accept the fifth alternative, namely, that value is that which can claim the approval of a majority in our own social group.[2] But this would seem to reduce ethics, aesthetics, and logic to a branch of statistical psychology based on counting votes—a kind of criticism by ballot box. And, since the standards of the majority are admittedly cruder than those of the philosophic few, we should be forced to agree with Bentham that, "prejudice apart, the game of pushpin is of equal value with the arts of poetry and music", and if we lived in a community where the masses revelled in watching a witch or a heretic suffering agonies at the stake, we should then be obliged to value what so plainly gained the general approval.

We thus seem to be thrown back on one or other of the first four assumptions. Now if we say that the type of approval we have in mind is that of 'persons of sound moral judgement', we shall be guilty of a *petitio principii*: for we shall then very naturally be asked, how are we to determine which particular persons possess a sound judgement, and the answer can only be those whose judgements enjoy the approval of ourselves or of their fellows. A factorist might perhaps suggest a correlational approach: we could invite a large and random sample of people to rank in order of goodness a series of typical actions from the most virtuous to the most vicious, correlate their rankings, and, if the correlations were predominantly positive, extract a general factor, namely, the average order of the whole set calculated by weighting the individual orders. We could then define 'a person with a sound moral judgement' as one whose ranking showed a sufficiently high correlation with the ideal order obtained for this factor. But that of itself could not serve to buttress up the subjective theory: indeed, if the correlations proved to be mainly positive (as in fact they do), the result would tend strongly to support the theory of objectivity. However, neither Ayer, nor any of the philosophers or psychologists who have followed him, have proposed a factorial solution: on the contrary, they seem rather to suppose that the correlations would for the most part approximate to zero, so that on their view the method would be inapplicable.

---

[1] *Introduction to Social Psychology* (1946), p. 124.

[2] This seems to be the form in which the theory is adopted by most behaviourists. Skinner, for example, writes "'You ought to love your neighbour' may be converted into the two statements: (1) 'Loving your neighbour is approved by the group of which you are a member'; (2) 'The approval of your fellow-men is *positively reinforcing to you*'. Both these statements may be demonstrated scientifically." (*Science and Human Behaviour*, 1956, p. 429; slightly abridged.)

Thus their final answer to our preliminary inquiry must be that they are talking of the judgements of each individual considered separately, with their own taken as typical. Ayer's actual wording is: " If *I* say to someone, 'You acted wrongly . . .', *I* am not making any statement; . . . *I* am simply evincing *my* moral disapproval." And presumably everyone is supposed to talk in the same personal way. It should be added that what he says of ethical judgements is intended to hold equally of aesthetic judgements, since, " in their case ", he says, " the same argument also applies ". What then is the argument?

If we turn to the original account of the theory as outlined by Ayer in the first edition of *Language, Truth and Logic* (to which Dr. Levi refers us), we learn that Ayer's main reason for putting forward an 'emotive theory of value' was that the objectivist doctrine " presents an insuperable objection to our radical empiricist theory ". Consequently he felt it was his " business to give an account of 'judgements of value' which would be both satisfactory in itself and consistent with our general empirical principles ". For the rest his own defence of the view is relatively brief.

So far as moral judgements are concerned the most detailed exposition of Ayer's view is that of C. L. Stevenson in his book *Ethics and Language* (1944). The general line he follows resembles that taken by Dr. Gregson and Mr. Whittaker; but his arguments are fuller and more systematic. He disowns any claim to *prove* the view that he advocates. He is content, he says, to explain it, and demonstrate that it is more plausible than any other. He begins, as Ayer did, by contending that, when people pronounce ethical judgements, they do not really mean to make assertions about definite facts. What they are doing is to communicate the emotional attitudes or feelings which they themselves experience. In a police court I could offer factual evidence to prove that John Doe stole my watch; but I could not offer factual evidence to prove that he acted wrongly when he did so. (The utilitarians who support a naturalistic view of ethical judgements would maintain that they *could* offer such evidence, but I should readily agree that their alleged demonstrations were invalid). Since therefore the assertion that stealing is wrong is not a proposition that I can verify by observational evidence or by any testable criterion, it follows, so Stevenson and his fellow positivists contend, that the assertion is meaningless. It expresses not an idea or a conclusion, but merely an emotion or feeling. This is in keeping with Professor Ayer's contention: " To say, 'You acted wrongly when you stole that money' does not state anything more than the concrete fact alleged, namely, 'You stole that money'. The words 'acted wrongly' are equivalent to an exclamation mark: it is as if I had said 'You stole that money' in a peculiar tone of horror." And doubtless the 'peculiar tone of horror' is intended at the same time to dissuade the pilferer from repeating his offence.[1] This interpretation, it is argued, furnishes an entirely adequate explanation of what are miscalled 'moral judgements'; hence there is no need to invoke any further assumption to the effect that " stealing is wrong " in a quasi-factual sense. The assumption of an ' objective world of values ' thus becomes a superfluous fantasy; and in virtue of Occam's maxim we may, and should, renounce it.

---

[1] Cf. Ayer, *Language, Truth and Logic*, p. 107.

Now the first stage in this argument is ostensibly a psychological account of what people experience and intend when they utter so-called moral judgements. Yet no concrete evidence is adduced to prove that this is in fact what they do experience or intend. The whole context makes it plain that the somewhat dogmatic description is derived, not from an actual inquiry into the experiences and intentions of different people, but merely from the general preconceptions of the author.

The champions of this ' emotive theory ' are careful to insist that, though sceptical about the existence of any objective standards, they are by no means sceptical about the importance of the subjective standards which common sense would commend, as their critics have so often suggested. On the contrary, Ayer himself believes that " those who accept the positivist analysis of moral judgements do not conduct themselves differently from those who reject it ".[1] The positivists indeed (so we are assured) would offer the same advice to others in problems of practical conduct as their theoretical opponents. " Ethical terms not only serve to express feeling; they are calculated to arouse it, and so to stimulate action." Thus the points of disagreement are not moral but only psychological, or (as Ayer seems to imply in his later writings) merely linguistic.

However, when he tells us that ethical statements are " calculated to stimulate action ", the positivist is surely going beyond the mere ' emotive ' interpretation which a strict positivism would require. He is implying that such statements are not descriptive but prescriptive: that they are in effect commands. Now, everyone who is not already pledged to the positivist view would undoubtedly agree that commands have a meaning, and can contradict one another. The law of Moses is " Thou shalt not kill "; the directive of Hitler is " You will now proceed to the final solution of the Jewish problem." The two statements are at once meaningful and mutually incompatible. Yet the positivist assures us that " only that which can be verified has a meaning ". And nothing in Ayer's discussion, so far as I can see, really succeeds in resolving the paradox which his views thus imply. In any case the construction which he thus places on all ethical commands does not dispose of such questions as—Is it *right* to kill or is it not? For the recipient of the command may still ask, why should I obey *this* command rather than that?

Moreover, the problem whether the daily conduct and the practical advice of the logical positivists are the same as those of the ordinary man is not itself the point at issue: the immediate question is whether their version of the ordinary man's experience and intention is correct or not. Now, as I shall try to show in a moment, it seems clear that, when the ordinary man declares it is wrong to steal, he does not mean *merely* to express his horror at theft: he sincerely believes that what he declares is a fact. If (say) a thief tried to maintain that stealing was not wrong, then, so the plain man would promptly reply, the thief's statement is false. Thus the positivist is *not* just giving us a more precise interpretation of what the ordinary man really does mean; he is telling us what the ordinary man *ought* to mean, or rather what he ought to say. The only difficulty is that, if the positivist is to be consistent, he himself must avoid " the meaningless word ' ought ' ". Strictly speaking, therefore, his own principles forbid him to claim that what he himself says can be a " factual statement which is true rather than false ". All that he can plausibly allege is that *he* as a philosopher 'feels horror' because ordinary men and women do not confine themselves to just asserting that *they* 'feel horror'.

---

[1] *Philosophical Essays*, pp. 245–7. He adds that the question whether or not the promulgation of such a theory encourages moral laxity is a matter for empirical inquiry, and I would suggest that it might form a useful problem for a statistical psychologist to investigate. From my comparative studies of delinquent and non-delinquent adults I should be inclined to say that, among those who consciously adopt the theory on more or less philosophical grounds, there is (as he says) no difference in standards of conduct, but that the popularization of such doctrines does indirectly tend to encourage moral laxity, and offers a mode of rationalization of which the more intelligent criminal is nowadays only too ready to take advantage.

The conclusion therefore must be, not merely that the subjectivists have failed to prove their case, but also that from its very nature they *cannot* prove it. To be valid, any argument which tries to deny that moral and aesthetic judgements are objective must be equally applicable to logical judgements. If the statement " X is wrong " merely means " I have a feeling of disapproval in regard to X, and would therefore recommend you not to commit X ", then the statement " Y is untrue " can only mean " I have a feeling of disapproval in regard to Y, and would therefore recommend you not to believe Y ". And the same sort of rejoinder can be turned against every other instance of the subjectivist theory of value.[1]

Finally, let us not forget that, since Professor Ayer first put forward his 'emotive theory' of ethics and aesthetics, now nearly a quarter of a century ago, he himself has appreciably modified his views—a change that Dr. Levi fails to notice. Ayer's present position is less extreme and less dogmatic. " To say, as I once did," he writes, " that these moral judgements are merely expressive of certain feelings, feelings of approval and disapproval, is an oversimplification . . . If someone wishes to say that ethical statements are statements of fact . . . he is welcome to do so; it is simply a question of how widely or how loosely we want to use the word 'fact'."[2] Thus, if I understand Professor Ayer aright, what he now wishes to do is simply (in his own phrase) " to recommend a new way of speaking, because [he] thinks that the old way is logically misleading or that [his] own proposal brings out certain points more clearly ".

III. COGNITIVE THEORIES

Cognitive Theories. Since the conative, the affective, and the emotive theories all seem to break down, we are left to assume that our ideas of value must be the result of some kind of cognitive experience. That need not of itself be incompatible with the view that value is subjective. Our ideas of colour and sound are also the results of cognitive experience; but many psychologists hold that these and other secondary qualities are purely subjective.

Cognitive processes are of very different types. Let us therefore begin by examining the chief forms which the attempt to construct a cognitive theory of value have taken in the past. There are two: the rational and the intuitive.

---

[1] Rather surprisingly a few supporters of the subjectivist view have in the past extended it in this fashion to cover the problem of truth. The best known instance of this 'logical subjectivism' is the doctrine commonly known as 'pragmatism', which for a while included among its champions at least one eminent psychologist, William James. " The theories we regard as true ", so James maintained, " are the theories that have won general approval . . . People believe in God because they find it *satisfactory* to believe in God." And for James this was not merely a statement of psychological fact or an expression of his own personal feelings; it became an exhortation. In *The Will to Believe* he assures us that the only test for the truth of a proposition is whether or not a belief in that proposition " works satisfactorily in the widest sense of the word ". The doctrine was avowedly an attempt to establish a theory of truth on a foundation of scepticism; and as such it is now almost universally acknowledged to have failed. I would therefore venture to contend that, as such an attempt does not succeed in the realm of logic, it must for the same reason fail in the realm of ethics and aesthetics.

[2] *Philosophical Essays* (1954), pp. 238f. In recent years there has been a marked tendency for both sides to come closer together. Ethical judgements, it is frequently argued, are not *mere* commands or exhortations, nor are they statements of fact in the ordinary sense of that word, but rather statements of " a queer sort of fact " (to borrow a phrase from Ayer). Cf. S. E. Toulmin, *The Place of Reason in Ethics* (1950) and R. M. Hare, *The Language of Morals* (1952), or (for the view of a philosopher who starts from Moore's assumption rather than from Ayer's) A. C. Ewing, ap. C. A. Mace, *British Philosophy in the Mid-Century* (1957, pp. 75f.).

(1) *Rationalism.* According to Socrates "virtue is knowledge"; and both Plato and Aristotle in their different ways tacitly assumed some kind of cognitive theory without explicitly working it out. The first systematic attempt to develop a theory along these lines came from the Cambridge Platonists.[1] In their view the moral life was the rational life. They adopted or adapted the Stoic motto: "*vivere convenienter naturae*". In the order of nature (so they maintained) there are certain immutable laws, both mechanical and moral, and these laws can be elicited by reasoning. Some of their later followers endeavoured to apply the same rational principles to the explanation of beauty, somewhat along the lines of the earlier Italian Neoplatonists (see below pp. 96f.); but the majority inclined to a more or less mystical interpretation.[2]

Now, if by 'reason' we mean the capacity to make inductive or deductive inferences no one can deny that, as thus defined, it may play a profoundly important part in moral conduct. Thus, as Ginsberg has so clearly shown, its practical function is to help us to discover what exactly are 'the facts of the case' in the widest sense of that phrase, i.e., not only to decide what is the precise nature of the situation and the problem it presents, but also to determine what will be the probable effects of different courses of action. Its theoretical function is to rationalize what seem to be our intuitive judgements: "for no particular intuition can claim finality; it is always subject to correction by further insight, and by being brought into relation with other intuitions".[3] But reason (as I have just defined it) cannot be the *sole* basis for our moral judgements. As Hume has insisted, from the proposition "these *are* nature's laws" we can never validly infer "therefore our behaviour *ought* to conform with them "[1]. Nevertheless, the inescapable conclusion is (and it is a conclusion which Hume himself failed to perceive) that, besides premisses which do not contain 'ought', *every* rational system of ethics must involve at least one intuitive premiss which *does* contain 'ought'.

---

[1] The leading member of the school was Ralph Cudworth, whose arguments are most fully set out in his *Treatise concerning Eternal and Immutable Morality* (1731, intended to be Part II of an unfinished work on *The True Intellectual System of the Universe*). His primary aim was to substitute a teleological system for the materialistic and mechanistic systems of Hobbes and Descartes. At the same time he desired to combat the notion, shared by Calvinists and Cartesians alike, that moral principles are merely a set of arbitrary commandments enjoined by the Deity. His essential contention is that moral values, like Plato's eternal ideas, are immutable parts of the 'intellectual system of the universe', and therefore to be ascertained not by observation but by reasoning. An exposition of much the same doctrine was developed by Samuel Clarke, the friend of Newton and Locke (*Natural Religion*, 1705, see esp. pp. 29–45). Spinoza's *Ethica ordine geometrico demonstrata* (1677) is a more formal effort at a rational demonstration, but based on a very different view of the universe. So far as our present purpose goes, Kant's doctrine of the moral reason might also be included with this group.

[2] The aesthetic theories of both Kant and Hegel are essentially intellectualistic and in some measure rationalistic. According to Kant, what gives us pleasure in apprehending a beautiful pattern is the adaptation of the form of the object to our cognitive faculties; but the pleasure follows the judgement, and "the validity of this pleasure is itself a law valid for all" (*Kritik der Urtheilskraft*, 1790, pt. i). According to Hegel "art is the presentation of Truth" (which for him means "spiritual reality") "in sensuous form" (*Aesthetik*, 1842). Both Kant and Hegel, it must be remembered, declined to recognize that nature could be beautiful.

[3] M. Ginsberg, *Reason and Unreason in Society*, 1947, ch. XIII 'The Function of Reason in Morals'. He puts his conclusion "in psychoanalytic terminology" by saying that "the function of reason is to make the *ego* increasingly independent of the *super-ego*". At the outset of his paper he appears almost to adopt the same view as Cudworth and followers. Moral action, he argues, is reasonable action; and, since "a distinctive mark of rational behaviour is purpose" (as distinct from unconscious impulse, habit, or social pressure and suggestion) "reasonable action is action for which a ground or reason can be given by reference to an end more or less consciously aimed at". I myself should qualify this by arguing (with Hume) that reason is concerned solely with relative ends, not with ultimate ends. But later in his paper Ginsberg remarks that "both sociological and psychological theories of ethics generally find themselves compelled, consciously or unconsciously, to reintroduce the *categories of value* which they have pretended to discard . . . In a great many cases the differences that divide individuals arise rather from differences as to the facts than as to their valuation".

2. *Intuitionism.* An intuitionist theory was developed by Shaftesbury and Hutcheson. "Should one who had the countenance of a gentleman", says my Lord Shaftesbury, "ask, why I would avoid being nasty, when nobody was present? I might give him a slight answer, and say: ''Twas because I had a nose' . . . And, even though I had neither nose nor eyes, my *sense* of the matter would be the same." In short our notions of what is nice or nasty, beautiful or ugly, right or wrong, are attributable to a kind of 'taste'.[2] Bishop Butler similarly holds that the dictates of morality are directly apprehended; but he maintains that the analogy with taste underrates their imperative character. Accordingly, for the designation 'moral sense' he prefers to substitute the term 'conscience'; and declares that "had it power as it has manifest authority, it would absolutely govern the world".[3]

Shaftesbury's "home-spun philosophy" was cast into a more articulate shape by Frances Hutcheson. The 'moral sense' he regards as part of man's equipment as a social being; and it is he, not Bentham, who first declared that the effect of moral action will be to bring about "the greatest happiness for the greatest number" (*Inquiry concerning Moral Good and Evil*, 1726, sect. 7). It is Hutcheson too who published the first systematic treatise on aesthetics (*Inquiry concerning Beauty, Order, Harmony and Design*, 1726: the two 'inquiries' originally formed parts of a single treatise). Beauty, he believes, like goodness, "is probably a simple idea which cannot be further explained"; and just as we have a 'sense' of humour, so we have a 'sense' of beauty. It is however a 'reflex sense', by which he means a mode of perception which implies the activity of the primary senses of sight and sound.

Hutcheson's theory, like that of Shaftesbury, and the Cambridge Platonists, was developed in opposition to the prevailing vogue of 'Hobbism'. A similar theory, in opposition to Hume, was subsequently elaborated, and couched in rather more scientific terms by Price, a dissenting minister and Fellow of the Royal Society, best known to statisticians for his two papers reporting and discussing 'Bayes' theorem'. In his work on 'morals' the problem he sets himself is to decide "whether right and wrong are real characters of action or only qualities of our minds", that is, in modern phraseology, whether they are objective or subjective. He finally agrees with Hutcheson in holding that moral ideas are "simple ideas", and therefore objective, and he defends his conclusion on psychological grounds, contending that such ideas must be ascribed to " some power of immediate perception in the human mind ".[1]

---

[1] This generalization, as a rule, is stated too broadly and accepted too uncritically. In my opinion the notion that one can *never* infer a practical injunction from factual premisses is mistaken. A puzzled Frenchman in a telephone kiosk was trying to ring up an old acquaintance. "What", he asked me, "*ought* I to do?" I replied in my halting French: "You desire a method of audible communication with your friend from this kiosk. In a London kiosk the method of audible communication is to press button A. Therefore, press button A". By the rules of traditional logic my conclusion was a hopeless *non-sequitur*. To make the syllogism valid it should be changed to "therefore you desire to press button A". But obviously he didn't; else I should not have needed to tell him to do just that. If however we substitute a more general logic of relations for the old syllogistic logic of subject-and-predicate, the deductive inference becomes perfectly valid.

[2] *Characteristicks of Men and Manners* (1711), pt. I, sect. iv. The author, the third Earl, was the grandson of the more celebrated (or more notorious) politician of Charles II's reign—Dryden's 'bloody bold Achitophel'; and the ancestor of the seventh Earl, who is commemorated by the statue of Eros in Piccadilly Circus. Locke, the friend of his grandfather, had been his tutor. His high flown sentiments and high falutin style are parodied by Berkeley in *Alciphron.*

[3] *Analogy of Religion, Natural and Revealed with the Course and Constitution of the World* (1733). The famous quotation is from *Sermon II.*

Price's ideas were taken up by Thomas Reid and fitted into the general scheme of his own philosophic theory. Reid had succeeded Adam Smith in the Chair of Moral Philosophy at Glasgow; and in his *Inquiry into the Human Mind on the Principles of Common Sense* (1764) he undertook the task of refuting Hume's doctrine that " the sole objects of immediate knowledge are sensory impressions ". It is in this connection that he introduces the novel distinction, which has since become a commonplace of psychology, between ' sensation ', whereby we " become aware of the mere colour and smell of a rose ", and ' perception ', whereby we " directly apprehend the rose itself as an object existing independently of our minds ". And he argues that we apprehend the beauty of the rose in the same direct and intuitive fashion as we apprehend the rose itself. Thus, " when a man pronounces a poem or a palace to be beautiful, he affirms something of that poem and that palace ". Similarly, " by an original power of the mind, when we come to years of understanding . . . we *perceive* certain things to be right and others to be wrong ".[2] Among modern writers the leading champion of the doctrine that we recognize what is good by immediate intuition has been G. E. Moore;[3] but it is perhaps too often forgotten how largely his realism, his intuitionism, and his reliance on ' the principles of common sense ' had already been anticipated by Price and Reid.

From this review it will be seen that none of the theories put forward by our critics and served up as the latest discoveries of modern behaviourism involves anything essentially new. Each has a long history behind it; and all the main alternatives seem now to have been covered, so that it appears doubtful whether any wholly fresh hypothesis is conceivable. Merely to change the wording or the incidental details is not to change the substance. *Nihil sub sole novum, nec valet quisquam dicere, Ecce hoc recens est.*

IV. An Empirical Investigation

Introspective Evidence. When we look back over all these various discussions, two defects cannot fail to strike us. In the first place, with a few notable exceptions, practically all the writers I have quoted are interested predominantly in one particular form of value—namely, ethical value. When they touch on problems of aesthetics, their references are for the most part incidental and perfunctory; yet nearly all acknowledge that the two problems are closely parallel. Secondly, the psychological evidence on which they rely is confined to their own somewhat biased introspections. Nevertheless, it is plainly the psychological evidence that constitutes the crux of the matter. During the last three or four decades, while most professional psychologists have fallen under the sway of behaviourism, introspective studies have been virtually taboo'd; and the few experimental inquiries that have been carried out have been limited to the collection of behaviouristic data, in the shape of preferential comparisons or rankings.

---

[1] *Review of the Principal Questions and Difficulties in Morals* (1758), pp. 11, 57. Price's chief opponent was Joseph Priestley, the chemist, who was also his closest friend: they published jointly *A Free Discussion of the Doctrine of Materialism and Philosophical Necessity* (1778). It was Price's sermon in support of the French Revolution that occasioned the publication of Burke's celebrated *Reflections.*

[2] Cf. Hamilton's edition of Reid's Works, pp. 492 and 598. Today Reid seems to be remembered chiefly as the originator of a somewhat elaborate form of the faculty doctrine, which was later popularized by Gall and the phrenologists.

[3] G. E. Moore, *Principia Ethica* (1903). Moore advances no formal argument to support his view, but rests it on the particular brand of philosophic realism that he derives from a general analysis of what he terms ' plain common sense '. From some later remarks in his ' Reply to My Critics ' (*Philosophy of G. E. Moore*, ed. P. A. Schilpp, 1942, p. 542) several writers have supposed that he eventually came to agree with C. L. Stevenson that the essential function of an ethical statement lies in its *emotive* meaning. But according to a paper published a few months ago by A. C. Ewing (*Mind*, LXXI, 1962, p. 251) Moore held to his old view—that such statements have *cognitive* meaning and are based on intuition—to the very last.

Over a long period of years, and with the help of several co-workers, I have myself endeavoured to collect introspective reports bearing on the foregoing issues from a wide and representative sample of persons, amounting by now to just over 200 men and women for most of the relevant questions. A synopsis of the results so far obtained is appended in a table at the end of this paper (pp. 103–4). The general method and the principal conclusions may be briefly summarized here.

I have found it instructive to divide my total sample into two main groups: first of all, a group which really represents only a very small selection of the total population (roughly about 4 per cent); they are all highly educated, and may conceivably be influenced by their specialized knowledge or theories; these I shall call the 'sophisticated' or 'S-group': and secondly all the others; these I shall refer to as the 'O-group', i.e., 'ordinary' persons. The classification was always made before the experimental results were obtained.

(1) I begin by showing my subject a couple of plain rectangular cards—one about the size of a visiting card (8 × 5 cm.), the other that of a postcard (13 × 8 cm.). And I ask (i) "Which is the larger?" (ii) "How do you know?" (iii) "Do you think the difference is there independently of your noticing it? Would it still exist if no one were in the room, or if all living creatures were wiped out?" (iv) "Does the difference imply a difference in some specific quality or characteristic possessed by the two objects, and if so what?" (v) "Does this underlying characteristic also exist independently of your noticing it?" The actual wording may be varied according to the vocabulary and intelligence of the subject I am questioning. The philosophical reader will note that questions iv and v relate to what Johnson has taught us to call the 'determinable' (*Logic*, Pt. I, chap. xi)—in this case size; whereas question iii relates to what he would call the 'determinates'—i.e. the particular sizes of the two cards: and similarly in the questions that follow.

(i) Everyone without exception indicates that the card measuring 13 × 8 cm. is the larger. (ii) Most say that they "see" the difference; some that the cards "look different": among the S-group several declare that they first "compare the two". At every level of intelligence there are a few who suspect some recondite trap, and instead of answering the second question directly, start talking round the subject. Quite a number proceed to demonstrate the difference by superimposing the smaller card on top of the larger; and I then ask, "What made you say that was the larger, before you moved them?" At this stage the main object of the discussion is merely to drive home that what is really wanted is a plain introspective report of first-hand experience (including thoughts and feelings as well as perceptions). (iii) Nearly all in the O-group maintain that the difference exists independently of the observer; in the S-group a few start arguing to the effect that "We don't even know that the cards exist when no one is there to see them". (iv) The majority spontaneously say that the difference is a difference in "size", though sometimes the answer is delayed by a certain amount of arguing or reflection to discover what the question means. Many in the S-group use the word "area". (v) Finally, practically everyone agrees that the size of the cards exists independently of observation, though a few in the S-group add—"providing one assumes that the cards themselves have an objective existence" (or words to that effect[1]).

(2) I then show the subject two cards of similar size—one a coloured reproduction of the face of the angel in the London version of the *Virgin of the Rocks*, the other a coloured and rather vulgar postcard of an ill-drawn and intentionally repulsive female face. With these I ask first (i) what difference the subject notices. Sooner or later nearly all say that the former is "more beautiful" or the latter "uglier". (The experiment is similar to the test of 'Comparing Faces' which Binet assigned to Age VI and Terman and

---

[1] In what follows, to avoid constant repetition, I shall assume that the reader supplies this qualification for himself.

I assign to Age IV.) I then continue with much the same questions as before. In both groups nearly everyone says that he " just sees " the difference. A few in the S-group explain that they find the former " more pleasing "; one or two say that it comes closer to their ideal, or that it is sexually more attractive, sometimes adding that the other is more amusing: these are the nearest approaches to any suggestion that the judgement was based on feeling rather than on direct perception, and they are exceptional. For the most part the replies with these two cards are given much more promptly than the replies to the corresponding questions about size. (iii–v). Practically all the members of the O-group say that they consider the *difference* to exist even when it is not being observed. A few find it more difficult to say whether the *beauty* existed when not observed. In the S-group there is a marked tendency to discuss the matter as a theoretical problem. The majority of students in the Arts Faculty (except philosophy students) assume that beauty does have some kind of independent existence as a quality of objects; just over half the students in the Science Faculty, and nearly all the students of philosophy, maintain that it does not. Among older men and women from the professional classes much the same divergence of opinion is apparent according as they have or have not obtained some kind of scientific or philosophic training.

(3) I next try to get away from judgements that are merely relative and objects which are man-made, and turn to the problem of natural beauty. For this purpose I use a bowl of flowers. I then say: " Look at that particular daffodil (or whatever flower is available). Would you agree that it is beautiful? What makes you say so? Do you see the shape of the flower as a flat pattern (as it would be, if drawn or painted on a sheet of paper), or in the round? Do you perceive the beauty of the flower in the same direct way that you perceive the yellow and the distinctive shape? Which caught your attention first—the beauty, the colour, or the shape? If you left the room, would the flower still be beautiful? Would it still be yellow? Would it still have the same shape? " These questions prove more difficult to answer with just a single word or two. They give rise to all kinds of doubts, reservations, and manifest theorizing. In the O-group the vast majority say they see the beauty in the same direct fashion as they see the colour and the shape, and almost as many assume that beauty, colour, and shape continue to exist even when unobserved. A few are doubtful about the colour existing when the room is in darkness, but apparently do not consider that darkness affects the shape. In the S-group a larger number incline to be non-committal. In the case of the flowers there are far more references to feelings and emotions—usually " pleasure ", " delight ", or (more rarely) " admiration "—than in the case of the pictures. Nearly half of those in the S-group doubt whether the beauty exists when there is no one to notice it. The beauty they associate for the most part with the colour; and they treat it as yet another kind of ' secondary ' quality. The shape, however, nearly all of them assume must remain the same whether observed or unobserved.

(4) For my last set of questions in the domain of aesthetics I start by inquiring what kinds of scenery the subject finds most beautiful—asking them to specify, separately wherever they can, some type of arctic, temperate, and tropical landscape. I then suggest that they should imagine two uninhabited planets—one a bleak and featureless bulk of cloud, the other displaying almost everywhere one or other of the types of scenery they have named as beautiful; and I inquire whether a solar system which included one or more planets of the latter kind would be in any way better than one which comprised nothing but planets of the first type, or whether, in the absence of observers, there would be no difference whatever in the aesthetic value of the two. This seems the most difficult problem of all. In general, most of the O-group end by giving much the same answers as in the case of the flowers. Of the S-group nearly a third say neither would be better nor worse than the other, unless an imaginary observer is tacitly assumed to be there to compare them; nearly a half think the inclusion of a planet with beautiful scenery would actually increase the value of the universe; the rest think the question meaningless or impossible to answer.

Oddly enough, the number of those who give affirmative answers to the question about independent existence is larger when it relates to *differences* in aesthetic value than when it relates to beauty in a single case (e.g., that of the flowers).

(5) Turning from aesthetic values to ethical, I proceed in much the same fashion. I begin with a comparison of two contrasted situations, and take a modernized version of the story of the ' Good Samaritan '. " Tiny Tim, who is blind, falls while walking down a narrow lane, and breaks his leg; Smith kicks him out of his way and walks on; Jones binds up his wounds, hails a taxi, takes him to hospital, and calls the next day with flowers to see how he is getting on. Which of the two actions is better? Why do you say so? If someone gave the opposite answer, would that in your view be an untrue statement? Have you any special feelings or emotions in regard to their conduct, and how far do these furnish the grounds for your judgement? Would one action still be better than the other if neither you nor anyone else knew about the incident—e.g., in a hundred years' time when all concerned were dead? " There is no need to discuss the answers in detail. The majority say that they immediately recognize, without stopping to think, that Jones's is better than Smith's and that if anyone gave the opposite answer, it would be " obviously untrue ". Many say they would also feel angry or indignant at Smith's action, but none says that this is the main reason for his judgement. In reply to the last question, a few of the more sophisticated say that Tim's pain would be increased by Smith's action and his comfort and happiness increased by Jones's, and consider this to be the sole ground for holding that Jones's action is better than Smith's: hence, they argue, when the whole incident is a thing of the past the question becomes meaningless, unless it is taken as an imaginary illustration or parable to guide behaviour should a similar situation arise now. But this, like most of the attempts at rationalization, nearly always sounds like an afterthought.

(6) Next, I ask them to consider the following situation. " Robinson Crusoe is the sole survivor of a nuclear war, and is himself slowly dying in agony.[1] How does he face his situation? (*a*) He is full of self-pity, cursing those who were responsible, and trying in vain to end his life. (*b*) He bears it in stolid silence, forgiving and excusing his enemies. Neither mode of behaviour can lessen his pain or alter his situation. Do you consider either of these attitudes better than the other, or is there nothing to choose between them? What makes you say so? What would you say if anyone tried to maintain the contrary? " Here the vast majority reply, with little or no hesitation, that the stoical attitude is the better, and that, if anyone maintained the contrary, he would be mistaken. All of them find considerable difficulty in explaining why they come to this conclusion. The commonest replies are to the effect that " patience or courage is finer than cowardice " and (particularly among the O-group) " everyone admires a hero " or " I personally despise a coward ": i.e., they explain their decision in terms of general or personal feelings. A number go on to suggest excuses for the coward—e.g., " the way he's probably been brought up ".

(7) When investigating juvenile delinquency it seemed desirable to undertake a study of what are sometimes called ' negative values ' or ' anti-values ', i.e., of judgements about what is evil or wrong in addition to judgements about what is good or right.[2] So far as our present discussion is concerned, the chief point of interest lies, not in the problem of evil in itself, but simply in the mode in which evil objects or actions are apprehended and in the assignment of objectivity or subjectivity. Here therefore I need refer to only a single set of answers. Amongst other questions I ask my subjects to picture one or two incidents involving the deliberate infliction of unmerited pain or suffering, e.g., a school bully who

---

[1] With the S-group the example I usually chose was that described by Hemingway in his well known story *The Snows of Kilimanjaro*.

[2] Cf. *The Young Delinquent*, p. 405 and refs. It was a little surprising to find that, among my delinquent subjects, even the habitual bullies and sadists nearly always replied, without hesitation (and in most instances, it seemed, with perfect sincerity), that the infliction of pain was wrong.

sticks pins under the finger nails of a younger boy, or a youth who wantonly destroys a small girl's cherished doll. In both cases the answers are very similar, and almost unanimous. None of our questions produced such prompt and confident replies. The number of those who declare that their judgements are based on an immediate and unreflecting recognition of the quality of act is larger than in any of the cases hitherto discussed; and nearly all agree that the wrongful quality of the act is independent of any human judgement. Among the S-group a small proportion go on to justify their judgements either by giving plausible reasons or by referring to personal feeling. One of them, himself a student of philosophy, remarks: "The pain felt by the small boy is not in the pin but either in his body or in his consciousness, and on either view subjective; the misery felt by the little girl is a conscious experience and nothing more: that too is therefore purely subjective. But in each case the situation itself is also evil, and its occurrence is certainly a fact independent of our recognition, so that here we plainly have a case where evil is objective." Another observes that "although I react to such situations with strong feelings of repugnance and displeasure, to say that these are the causes of the wrongfulness would be like saying that opening one's umbrella is the cause of rain".

(8) Finally, I ask each of my subjects what is the most impressive experience (of beauty or the like) that he has ever had—particularly whether he has ever felt himself to be in the presence of what Burke and others call 'the sublime', or—what in my opinion are closely related to it—the 'numinous' or the 'mystical'.[1] Such experiences are seldom alluded to in the stock disquisitions on 'value'; and the few writers who have discussed or investigated them seem largely to have ignored the sense of meaning and significance which they so often convey. Nevertheless, if I may trust the replies to my questions,

---

[1] A word is perhaps desirable as to Kant's distinction between the sublime and the beautiful. Although he is reluctant to allow that Nature or natural objects may be beautiful, he maintains that they are often sublime. However, few of my own subjects draw a contrast between the two; rather they seem to think of the sublime as a species of beauty. They commonly deny that their own experience included terror, horror, or despair, and hence they would certainly reject Kant's earlier account of the sublime as involving something hostile and *Schreckhaft* (a view previously put forward by Burke in his *Inquiry into the Origin of our Ideas of the Sublime and Beautiful*, I, vii, II, ii). The word that most frequently occurs in the reports of my subjects is 'awe'. So far from being depressing, the ultimate effect is said to be exhilarating and optimistic—a kind of *catharsis* similar to that experienced when the curtain goes down on a Shakespearian tragedy. Beethoven held a similar view of music: "Wem sich meine Musik verständlich macht, der ist über allen Jammer der Welt erhaben". Many would endorse Kant's later account: "so giebt auch die Unwiderstehlichkeit ihrer Macht" (i.e. Nature's) "uns als Naturwesen betrachtet, zwar unsere physische Ohnmacht zu erkennen, aber entdeckt zugleich ein Vermögen, uns als von ihr unabhängig zu beurtheilen, und eine Ueberlegenheit über die Natur" (because we are conscious and rational beings whose thoughts can transcend the physical limits of Nature: *Kritik der Urtheilskraft*, sect. 28).

In support of my claim that mystical experiences and aesthetical are closely allied, may I quote the most eminent of British experimental psychologists, Charles Myers? After describing a series of experiments on musical appreciation carried out in the Cambridge laboratory, he concludes: "There are three lines of activity which take us away from practical preoccupations: the simplest is play; the second, fantasy or daydreaming; the third consists in mystical experience, in which, as in love, we lose awareness of our own individuality and its relations to our surroundings. I believe that our experience of beauty always partakes in some degree of this ecstatic character. Nowhere do we feel this 'uplifting' so keenly as in music" (*ap.* Max Schoen, *The Effects of Music*, pp. 36–7).

The English language has no word to describe the objects or situations which arouse all such experiences: the word 'sublime' is too narrow; the word 'holy' has been spoilt by theological associations; Otto's 'numinous', with its echoes of 'ominous', is perhaps more appropriate; Aeschylus's δεινός (Liddell and Scott, sense II) comes near to what is wanted; but perhaps best of all is Isaiah's triple 'qadosh'.

persons who claim to have had such experiences almost always ascribe to them (or to the objects so experienced) the highest value of all.

Quite unexpectedly, this turned out to be by far the richest of all the fields of inquiry to which these ‘ axiological studies ’ have led. Here only the briefest summary is possible. Taken as they stand, the figures obtained indicate that experiences of this type are much more frequent than is generally supposed. At first sight they would seem to be commoner among the S-group than among the O-group; but that may be due partly to the fact that persons in the O-group more rarely find themselves in situations that are apt to trigger off such experiences,[1] partly to their inability to recognize or put their experiences into

---

[1] This is intelligible when we note the type of situations usually mentioned, e.g., ‘ the pyramid of Gizeh ’, ‘ the Victoria Falls ’, ‘ a volcano in eruption ’, ‘ on first entering St. Peter’s ’, ‘ the first sight of the Himalayan peaks ’, ‘ a thunderstorm at sea ’, ‘ sunrise in the Alps ’, ‘ the Holy Places ’ (in Palestine), ‘ the last movement of the C Minor symphony ’, ‘ the *Liebestod* in *Tristan* ’ ‘ Prometheus’s prophetic speech at the end of Act I ’ (the reference was to the lines beginning ‘ Remit the anguish of that lighted stare ’ in Shelley’s *Prometheus Unbound*) The bookish types refer to experiences while reading certain passages in Wordsworth, De Quincey, Sir Thomas Browne, or the Bible (usually Job, Isaiah, or Ezekiel). In the ‘ ordinary ’ group the situations mentioned were of a more ordinary kind—‘ a starlight night ’, ‘ a glorious sunset ’, ‘ a rainbow ’, ‘ the silence on the top of a mountain ’, ‘ waves breaking over a rocky coast’, (in each case the time and place being specified or named), caves, glades in the wood, eerie places such as churchyards or ruins on a dark night or by moonlight, being lost or alone, funerals, deaths and rooms, houses or places associated with death, religious services, being in love, being one of an excited crowd, and (with older men) various war-time situations.

One point emerged which is overlooked in studies of mystical experiences based on literary sources. Many of my subjects, particularly those in the O-group, reported a kind of ‘ fellowship-mysticism ’, quite distinct from the more usual ‘ nature mysticism ’ or ‘ soul mysticism ’, which usually arise in solitude. The clearest accounts came from the Welsh; the experiences described occurred, not only during religious meetings (particularly those of a revivalist type), but also on more ordinary occasions, such as community-singing and even discussion-parties.

A small proportion in both groups (chiefly women) related experiences ostensibly of an ‘ extra-sensory ’ type, often described in phrases borrowed from some theosophical or spiritualistic creed. This commonly made it difficult to distinguish between the actual nature of the experience and the interpretations placed upon it. Most of these I consequently excluded. One report is worth noting for its bearing on the connection between two types mentioned above. A solitary student, who, like Richard Jefferies on the hill top, had often experienced a kind of semi-mystical “ communion with Nature ”, heard a voice saying: “ Love not Nature, love your neighbour ”, and changed to the more social type: (as an audile, he believed the voice was simply auditory imagery provoked by his own subconscious misgivings hitherto repressed). None of those I questioned had been the subject of a ‘ mystical conversion ’ of the type described by William James (*Varieties of Religious Experience*, pp. 189f.). The nearest instances were the reports of three or four self-confessed sceptics who recorded an experience like those so vividly described by Browning’s intellectual cynic: “ Just when we’re safest there’s a sunset touch, a fancy from a flower bell, someone’s death . . .” (*Bishop Blougram’s Apology*).

Nor had any of the members in my groups taken any of the drugs (mescalin or the like) now so often used experimentally to produce quasi-ecstatic states. While under gas at the dentist’s, one had experienced something like the “ rapturous sense of a reconciliation of self and other, of good and evil, followed by a revulsion of mood to horror ”, which William James tells us he felt during inhalations of nitrous oxide. Another, while taking regular doses of morphia, had (as he put it) “ been wafted off to a delicious Never-never-land ”—apparently a sort of Lotus-eaters’ paradise—which, he maintained, had a “ genuine reality or value of its own ”, and was “ not as illusory as you might suppose ”.

When psychoanalysed (in the way one would psychoanalyse an ordinary nocturnal dream), the contents showed a close connection with the individual’s waking fantasies and daydreams, and with his suppressed emotional interests or conflicts. Thus the drug seemed to release rather than to produce the content of the experience. And it was equally clear that suggestion played an appreciable part in determining both the actual details and what I suspected were ‘ secondary elaborations ’. In any case, as all these records abundantly show, there exists a vast mass of unexamined words, and partly to a fear that they might be thought either "slightly pathological" or "unduly superstitious" or (as a Scottish girl put it) "a wee bit fey". I found, as Galton did, that at the beginning of the interview, the person questioned would often return a slightly embarrassed denial, but, after his reticence had been overcome, he would later confess: "Well, there was one curious experience . . ."

The introspective accounts given of this group of experiences make them far more difficult to assess than those discussed in the previous sections. The language used, though still predominantly cognitive, nearly always implies affective or emotional ingredients, and often what may be called ethical or conative ingredients. Such experiences therefore seem to be highly complex. A comparative analysis suggests that they may include at least three, and usually more, of the following. (i) First and commonest of all, an impression of *super-human power*, or of immense size[1] suggestive of superhuman power, sometimes inducing in the beholder an almost mesmeric fascination. Then as a primary reaction to this there is nearly always (ii) a simultaneous realization of *human impotence*, or rather of a compulsory submission or obedience to an 'ineluctabile fatum' which the manifestation of such a power implies. Yet at the same time there is often (iii) a growing impression that the power is, in some inscrutable manner, *purposive and law-abiding*, and so not merely a physical power but a moral power, or (as one of my subjects puts it) "a power which in ways unknown is both just and friendly". As a further consequence there supervenes (iv) an impression of value triumphant, 'a harmony so perfect that it transcends human expression',[2] which imparts a kind of reflected moral power, and often leaves behind it a more lasting mood of *uplift or exhilaration*. Throughout there is generally (v) a feeling of complete *temporary detachment* from the practical problems and personal interests of everyday life, so that both the experience itself and what is experienced is often described as timeless: and finally there may also be (vi) a transitory sense of *union or coalescence* with the power that has been thus manifested (or some more permanent power behind that power), often through a kind of *Einfühlung*, or wish to "share the impulse of its strength".[3]

---

data still awaiting systematic and critical analysis; and, to my mind, the experiences occurring more or less naturally deserve quite as much attention as those artificially produced by drugs or other devices.

[1] Only one of my subjects found "the infinitely small as impressive as the infinitely large —what I see in my microscope as sublime in its way as what others see through a telescope". In a nuclear age this seems rather surprising. After all, gamma-rays are more powerful than sheet-lightning, a virus more deadly than an elephant; and the fact that there are over 30,000 million million elementary particles in a grain of sand is surely more amazing than the estimate of 30,000 million stars in the galaxy. The psychologist has little difficulty in guessing at the probable reasons; and yet the lively discussion which this and similar anomalies so readily provoke suggests that there is here a further problem well worth investigating.

[2] John Buchan's way of describing an 'hour of revelation' he himself once experienced: (*Memory Hold-the-Door*, p. 49).

[3] Shelley, *Ode to the West Wind*; (cf. also his *Hymn to Intellectual Beauty*). When I ask my subjects about their own emotional attitude during this final state—whether it is one of pleasure or pain, desire or dread or approbation in its simple or more complex forms (admiration, veneration, or the like), or possibly a kind of anaesthesic Nirvana, the phraseology which, rather surprisingly, appears again and again is (to quote one of them): "It's like *being in love*, which somehow combines them all, and the love seems mutual—though that, you'll say, is the 'pathetic fallacy'". They quote Wordsworth, or "Stopford Brooke's poem about 'Nature's Love for the Nature-lover'", or Richard Jefferies' *Story of My Heart*.

This identification of self with Reality or Nature is usually supposed to be characteristic of Indian rather than of Western thought; the recurrent formula, *Tat tvam asi*, crops up first in the Vedanta, then in the Ghita and the cult of Atman, and finally in the jargon of the theosophical trance-talker. The Sufi poets of Persia make use of similar philosophical flights, and were doubtless influenced by the Indian mystics as well as by those of Oriental Christianity and Neoplatonism. But they lay far greater stress on experiences of cosmic beauty, which they do not hesitate to compare

The reader familiar with previous discussions of this and allied problems will notice that to some extent the points in the foregoing empirical list correspond with the specific characteristics included in Bradley's 'four main features of experiences of the sublime' and with those mentioned by Rudolf Otto as the distinguishing marks of what he calls 'the sense of the numinous'. In their details many of the accounts—and not only those which could fairly be taken as referring to mystical states—agree with the various reports analysed by Professor Stace and Miss Laski in their recent studies of mysticism and ecstasy.[1] I may add that, among the persons in my O-group who recorded experiences of the type I have tried to describe, the majority seem to have felt that they were (as one of them puts it) "discovering something about reality", or at least about a part of it. Of the persons in the S-group just over half say so.[2] Yet even the few who consider such experiences to be essentially 'subjective' nearly always insist that they have a 'value'.[3]

General Conclusions. From all these various results two conclusions seem to stand out: first, the vast majority of those whom I questioned regard both aesthetic and ethical judgements as *assertions of fact*; value and differences in value they plainly think of as things existing independently of human feelings or human recognition and in that sense as genuinely 'objective'; secondly, most of them also maintain that their judgements are based on a *direct apprehension* of the valuable qualities and of the different degrees in which they are manifested. However, the mere discovery that a majority of persons in a random or representative sample hold that something or other is directly perceived and must therefore be an authentic quality of the external object, cannot of itself form conclusive proof that their interpretation is actually true: again and again in the past things that seemed self-evident have turned out on closer scrutiny to be undoubtedly false. Nevertheless, when the results of an experimental inquiry, as well as the general voice of common sense in almost every age and community, agree in supporting the objectivity of value, we are clearly faced with a strong presumption in favour of that conclusion; and the *onus confutandi* is thrown squarely on the shoulders of those critics who expect us to abandon an interpretation that appears so plausible.

V. OBJECTIONS AND REPLIES

What objections then are left to such a theory? Dr. Levi has told us "there are two that are commonly considered fatal: first, the frequent disagreement between individuals in specific cases, and secondly, the fact that the aesthetic quality of an object is itself dependent on other qualities that are acknowledged to be merely secondary and subjective—the beauty of a picture on the colour, and of a piece of music on the sounds: and similarly with ethical values".

---

with states of intoxication or ecstatic love. Nevertheless, we find the same doctrine turning up in the Western mystics and our own lyrical poets. Byron, to take only a comparatively late example, frequently refers to a transient sense of union:

> Are not the mountains, waves, and winds a part of me
> And of my soul, and I of them?

[1] A. C. Bradley, *Oxford Lectures on Poetry* (1909), R. Otto, *Das Heilige* (1917), W. T. Stace, *Mysticism and Philosophy* (1961), Marghanita Laski, *Ecstasy* (1961). I am also much indebted to F. L. Lucas, *Literature and Psychology* (1951): see especially the chapter on 'Values'.

[2] We might, I fancy, fairly apply to the experiences reported by the members of my own groups what Professor Price says of the mystical experiences reviewed by Professor Stace; "if we think them 'subjective' we have to admit that their subjectivity resembles the subjectivity of *perceptual* illusions or hallucinations, not the subjectivity of *emotional* states" (though, of course, latent emotional factors might be largely responsible either for their nature or even for their occurrence): cf. also Thouless's remarks about the experiences of contemplative mystics in his *Introduction to the Psychology of Religion*, pp. 226f.

[3] For brevity I have employed these semi-technical terms to summarize the gist of the replies; it is hardly necessary to quote them *verbatim.*

(i) *Individual Differences.* "In all other fields", it is argued (p. 51), "we adopt a simple criterion. When a perceptible property varies only with variations in the different objects and not with variations in the different perceivers or the different conditions under which the property is perceived, then we regard it as objective; on the other hand, when for one and the same object it varies with the perceivers or the conditions, then we regard it as subjective. In music, poetry, and painting it is notorious that what one man hails as beautiful, another dismisses as detestable. Similarly in morals: the Church still considers perfect chastity the ideal; *Sixth Form Opinion*[1] recommends premarital intercourse as a prudent preliminary to formal marriage. Beauty and moral goodness in all their forms are therefore properties which unquestionably must belong to the category of the subjective."

Now the premisses on which this twofold syllogistic inference is based—premisses which are implied rather than explicitly stated—are decidedly sweeping. In general what varies with the perceivers are not the properties themselves (as the argument requires), but the *assessments* of the properties. In physical experiments the scientist's pointer-readings nearly always "vary with the perceivers and the conditions under which they are perceived"; but we do not therefore infer that what is measured is subjective. Hence the crucial question is, *how widely* do the assessments fluctuate? And Miss Williams very wisely asks: "What is the evidence for these 'notorious' differences in regard to beauty and moral goodness? If" (she continues) "the phrase 'one man' refers—as the context suggests—not to the specialist or the expert, but to the ordinary man-in-the-street, why should we worry about what he thinks of this or that painter, poet, or musician, any more than we worry over his views about flying saucers or seek his opinion about assumptions of Euclid or Einstein regarding the nature of space? Surely it is the verdict of the expert and the specialist that should count."

However, in the field of aesthetics it is by no means difficult to find instances in which the pronouncements of experts and of connoisseurs have been wildly at variance. Landor thought the *Iliad* "mostly trash"; T. E. Lawrence, after translating the *Odyssey*, decided it was a "poor poem and a creeping piece of work". To Horace Walpole Dante seemed "a Methodist parson in Bedlam". Wordsworth insisted that Byron's verse was "not English"; and Byron retorted that Wordsworth was "the blind monarch of the one-eyed". Matthew Arnold calls Shelley an "ineffectual angel"; and Lamb described Shelley's poetry as "an obnoxious squeak". Macaulay declared that there were "no compositions in the world which approach nearer to perfection than the novels of Jane Austen"; George Eliot concurred. On the other hand, Charlotte Brontë set her down as "a very incomplete and insensible woman"; and Professor Garrod describes her rural scenery as "a land flowing with milk and water". One and the same literary critic may even disagree with himself. Not long ago we were told that our illusions about the merits of *Paradise Lost* had been finally dispelled by T. S. Eliot; but Mr. Eliot has since recanted, and Milton now stands reprieved. As for sculpture and painting, we find Shelley himself assuring us that Michael Angelo's statues are "wanting in dignity" and the frescoes in the Sistine Chapel "deficient in majesty". And, in our own day, the outcry against modern art has only been equalled by the abuse heaped on modern music.

But, the older Disraeli long ago observed, these varying verdicts are merely some of the 'curiosities of literature' and of the art-criticism of the day. Were a famous theatrical critic to announce, in the words of the schoolboy's textbook, that "Shakespeare is our greatest dramatist", that would "not be news"; but when Shaw startles us by declaring that "with the exception of Homer, there is no eminent writer I despise so entirely as Shakespeare",[2] every newspaper prints it. Each of our younger pundits (and this is as true in psychology as it is in aesthetics) would feel that he was lacking in originality if he merely continued to echo like a parrot the sentiments of the preceding generation. Moreover, even the finest of critics may have his momentary lapses; and, like the fairy queen, fall into an ecstasy over an ass's head. Merely to quote a string of isolated instances, some perhaps deliberately designed to sound paradoxical and pungent, can prove little or nothing.

---

[1] *Sixth Form Opinion* is a magazine edited by senior pupils in sixth forms; the reference is to the issue for 18 February, 1963.

[2] *Dramatic Opinions and Essays* (1907), II, p. 52.

The serious investigator must therefore beware of the common temptation to magnify the importance and the frequency of exceptional deviations, while ignoring the numerous humdrum examples of general agreement. His business is to take a large and representative sample of the total population, and check the hasty pronouncements of the impressionists and the iconoclasts by an impartial statistical comparison.

Let us therefore turn once again to the results of actual research. In collaboration with Miss Margaret Bulley, Dr. Dewar, and a number of art teachers and expert critics, I have presented reproductions of a wide variety of pictures, and photographs of buildings, statues, household objects, and natural scenery to a large group of persons so selected as to represent, in due proportion, all sections of the population—'experts and specialists', as well as 'the man-in-the-street'. Often the Royal Academician would warn us in advance that his order would completely reverse that of the modernist; and both were convinced that they would disagree with the judgements of the ordinary man or woman. Yet, surprisingly enough, when actual rankings were obtained, the correlations were all positive; and the averages in the various inquiries never fell below 0·70 and sometimes rose to over 0·85. The Academician might place a winter scene by Breughel 6th or 7th in a series of 60, and Klee's *Image d'hiver* 15th or 16th; the modernist might invert the order. But each would rank both of them far above the frosted landscape of a cheap Christmas card or the pin-up portrait of a popular film star; and, save for a few characteristic exceptions, even the most uncultured members in our group preserved the same *relative* distinctions. In similar experiments on the appreciation of music and poetry the correlations were again invariably positive, though not quite so high (0·60 to 0·75). Thus in all our tests the minor differences between individuals were entirely swamped by the general agreement.[1] These results plainly corroborate the view expressed by Burke at the outset of his *Inquiry*. "It is probable", he writes, "that the standard both of reason and of taste is much the same in all human creatures . . . And, indeed, one may observe that there is rather less difference upon matters of taste than upon most of those which depend on naked reason: men are far better agreed on the excellence of a description in Virgil than on the truth or falsehood of a theory of Aristotle."[2]

In the case of moral judgements it is the differences in standard between different countries, different epochs, and different civilizations which have supplied the chief ground for the ethical sceptic rather than the differences between individuals living within the same community. The oldest of historians relates that Darius, King of Persia, asked certain Greeks how big a bribe they would take to eat their dead parents. "Not for your kingdom", they replied. Then, turning to certain Indians, who considered it a pious duty to eat the bodies of their parents when they had died, the king inquired what they would take to burn the bodies like the Greeks. " Not for all the world ", was their horrified answer. " So powerful ", observes Herodotus, " is tradition and custom." And his view is echoed by many modern moralists. " There is ", says Canon Rashdall, " hardly a sin or vice according to our own notions which has not, in some circumstance or at some date, been looked on as a moral or religious duty " (*The Theory of Good and Evil*, p. 17).

---

[1] For a fuller account of methods and results the reader may turn to the summary given in Valentine's *Experimental Study of Beauty* (1962, pp. 146f.) and the references there cited. I may add that more recently Dr. Aydin Cancardas, using the same set of reproductions, has secured rankings from persons of different nationality and different cultural backgrounds, and still finds comparatively high correlations.

[2] *Op. cit.*, 'Introduction: On Taste'. Burke rightly indicates the two main causes of such individual differences as are observable. First of all, individuals differ by nature in their sensitivity to beauty, just as they differ in all their other perceptual capacities; and, since the objects or situations which occasion judgements of beauty are far more complex than those in which we are called upon to make judgements about size, warmth, or weight, the variations are wider, and the capacity for discrimination needs more preliminary exercise. The second cause arises out of the wide differences in intellectual equipment, opportunities for experience, and interest in the subject-matter concerned.

" Half the crimes of Clapham are chaste in Khatmandu."

But these and similar generalizations, though reiterated time after time, are by no means borne out by first-hand studies of the actual facts. If we turn to any systematic and impartial survey—that of Westermarck, for example, who can hardly be suspected of bias in favour of an objective theory—we learn that " all the major virtues have been acknowledged and practised by every people at every time, imperfectly to be sure, but in some degree beyond all doubt " (*The Origin and Development of Moral Ideas*, 1917). Admittedly there has been a vast if somewhat vacillating progress since man first emerged from barbarism,[1] with countless set-backs and reversals; but that holds true in almost every other domain. However, even if people varied as widely as is claimed, this would scarcely prove the conclusion. It would merely demonstrate that men's ethical judgements and standards of conduct are nearly always determined, to a greater or a lesser degree, by non-ethical considerations—particularly by the views and conventions prevailing in the society or in the group in which they have been reared, just as their intellectual judgements and their standards of reasoning are apt to be affected by irrational and non-intellectual influences, partly personal, partly social. Almost invariably the situations which call for a considered moral judgement are highly complex, and a certain degree of innate intelligence, educational training, and moral discipline is needed to carry out the requisite preliminary analysis and to overcome personal prejudice.

The experimental evidence is here far more meagre than in the case of aesthetic judgements. In my studies of delinquency I adopted much the same general method as in the earlier investigations on artistic appreciation. I collected a series of short descriptions (*a*) of typical actions—virtuous, vicious, and criminal—and (*b*) of imaginary persons regularly performing such actions, and asked the examinee to rank them in order of ' goodness ' or ' wickedness ' (using whatever terms seemed most appropriate to the individual tested). The tests were used with large and representative samples of children and adults, both delinquent and non-delinquent. As before, the correlations were all positive, ranging from 0·55 to just over 0·70.[2] The argument from individual differences therefore seems plainly to fall to the ground.

---

[1] Mr. Whittaker, who cites Darwin and Huxley in his support, tells us that " moral progress and the high moral standards on which we plume ourselves today are based, not on the ' eternal verities ', but on the results of the struggle for survival between different tribes, nations, social classes, and religious sects: they are biological utilities resulting from biological evolution, not ethical values resulting from increased moral insight ". Yet this is not the conclusion drawn by the evolutionary biologists themselves. " Nature ", says T. H. Huxley, " is no school of virtue, but the enemy of ethics. Since law and morals put restraints on the struggle for existence between men in society, the ethical process is in opposition to the cosmic process " (Romanes Lecture). And Darwin, in less rhetorical language, repeatedly stresses the same contrast.

[2] See *The Young Delinquent*, pp. 405f. for these and other tests, and for references to similar inquiries. Mr. Whittaker complains that, in discussing the treatment of delinquent children, I endeavour to reimpose the " outdated ethical code of the Victorian moralists ", which, he says, I apparently regard as " absolute and universal ". But my concern was with conformity to the legal code rather than to any particular moral code. I should be the last to set up Mrs. Grundy as the final arbiter of morals for the 20th century youngster. No code is sacrosanct. As I have argued elsewhere, ' goodness ', like ' beauty ', is not a matter of cut-and-dried rules and regulations. The

(ii) *The Subjectivity of Secondary Qualities.* The second of the two arguments that are commonly held to demolish the objectivity of value appears most frequently in popular discussions of *aesthetic* value, though with the necessary modifications it is occasionally carried over to the analysis of *ethical* judgements as well. " The appreciation of a picture ", we are told, " depends essentially on the pattern and harmonies of the colours; the appreciation of a piece of music on the pattern and harmonies of the sounds; and, since colours and sounds are wholly subjective—mental contents rather than attributes of objects, it follows that their beauty must also be subjective." Beauty in fact " is altogether in the eye of the beholder ".[1]

This is just a modernized version of the old strip tease performance, whereby all the perceptible qualities of the concrete object are peeled off one by one until we are left with nothing but a bare skeleton of spatial or numerical properties. But as a refutation of the realist position the argument is far from conclusive. To begin with, many of those who hold that beauty is objective, G. E. Moore for example, also hold that so-called ' secondary qualities ' are objective: after all there is no empirical evidence to prove that a material object, which the physicist would treat as consisting essentially of ' primary qualities ' (i.e., those that lend themselves to measurement and mathematical arguments), cannot likewise have the secondary qualities attributed to it by the naïve realist.[2] But secondly, even if the surfaces of objects possess no actual colour, and even if the air contains no actual sounds, the physical stimuli have attributes whose differences and variations run almost exactly parallel with the differences and variations in the colours and sounds which human beings perceive. In music much of the theory of melody, harmony, rhythm, and key-relations, and of the consonance and dissonance of chords, has in fact been based on this correspondence. And similar theories have been advanced by painters and art-critics in their endeavours to interpret the traditional notions of the ' harmony of colours ', of ' mixed ' and ' pure ' tints, and the like. Dr. Gregson, however, as we have seen, connects our impressions of aesthetic ' value ' not directly with the colours or sounds themselves, but with the ' hedonic tone ' which is found ' on introspection to pervade them '. Such a ' hedonic tone ', I readily admit, must be ' subjective ' (in Dr. Gregson's sense of the word)—as ' subjective ' as the ' positive hedonic tone ' we experience in the odour of a rose or the flavour of a vintage port, or the ' negative hedonic tone ' of a sting from a nettle or a burn from a red-hot poker—which even the most naïve of realists would not ascribe to the material objects. But I doubt whether mere sensory pleasure is an *essential* element in beauty as such. Dr. Gregson tells us that it " indicates that valuing is taking place ". If so I should regard the ' values ' thus attested as instrumental rather than as genuinely intrinsic.[3] Beauty depends not so much on the *qualitative* characteristics of the object as on its *relational* properties—on what Clive Bell in an oft-quoted phrase called ' significant form ', i.e.on its

---

only absolute commandment says neither ' Do not ' nor yet ' Do ', but ' Be ': ' Be ye therefore perfect (τέλειοι) '—or, as Aristotle puts, it ἡ ἀρετὴ ἐστι τελείωσίς τις (*Phys.*, VII, 3, 5; *Metaph.*, IV, 16, 3). The principle that " a thing is ' good ' in proportion as it approximates to an ideal and perfectly efficient specimen of its kind " applies to human beings as to all else. In the long run therefore to reform a delinquent is to re-form his character, not to condition him to a set of conventional rules.

[1] The more usual form of this familiar dictum (without the adverb ' altogether ') comes, not from Hume, as Dr. Levi supposes, but from Margaret Hungerford's novel *Molly Bawn* (1878)—a best seller in its day. It is there given as a quotation; and the allusion doubtless is to the earlier remark, worded as I have given it in the text, which is to be found in Lew Wallace's *Prince of India* (1867, III, vi, p. 78).

[2] See ' The Sense Datum Theory ', this *Journal*, XV, pp. 190 f.

[3] When I read the score of Bruch's G Minor concerto, and hear in my mind's ear the solo melody in the *Adagio*, I doubt if any ' hedonic tone ' attaches to the *sound* of my mental violin, and yet I enjoy the music almost as much as if I were listening to it at the Festival Hall. On the other hand, if I heard it broadcast by an ill-tuned receiver, the whole thing would be ruined, as surely as a water-colour sketch by Turner himself would be ruined if painted with a dirty box of paints.

structural properties as directly perceived.[1] It is chiefly for this reason that we hear so much, in the experimental studies referred to by Professor Valentine and Mr. Whittaker, about such characteristics as 'symmetry', 'balance', or 'due proportions', 'golden sections', and the like.[2] If, therefore, as some have held, we could *define* 'aesthetic value' in terms of certain specifiable relational properties, there would be no need to regard value as an irreducible concept. And a similar interpretation might be found for ethical value—possibly along the lines of the earlier rationalists. Let us therefore consider for a moment what evidence there is in support of such a view.

The Objectivity of Relational Properties. Asked what exactly do they perceive when they see an object that they recognize as beautiful, a number of persons in my S-group replied with some such phrase as 'harmony', 'variety organized in unity', or its equivalent. It is a notion which, as most of them agree, they have probably picked up from previous discussions or from books. And indeed both the theory and the phrases have a fairly long history behind them. They go back through Plato and the Neo-platonists to Pythagoras and his famous doctrine of 'attunement'.[3] The principle that simple ratios should hold between related parts is a canon that regulates not only Renaissance architecture but also the structural style of much Renaissance painting, sculpture, and music. It starts, at least as a systematic theory, with Alberti. His guiding maxim was that a true work of art should be so designed " ut sit pulchritudo quidem certa cum ratione concinnitas universarum partium in eo cuius sint: ita ut addi, aut diminui, aut immutari possit nihil, quam improbabilius reddat ".[1] Ghiberti puts it in a single phrase: " la proportionalità solamente fa la pulcritudine ". Music, it was said, is " geometry translated into sound: architecture is music translated into geometry ": both are reflections of the harmonic proportions which give order to the Cosmos.

---

[1] Clive Bell, *Art* (1914). Roger Fry in his 'Essay in Aesthetics' (*Vision and Design*, pp. 30f.) describes "true beauty" as consisting in "a perception of purposeful order", and thus in its purer manifestations as "supersensual". Such a perception requires a kind of 'distancing'—a temporary detachment from our usual interest in the 'sensual' or concrete aspects which commonly determine practical utility. In ordinary life our work-a-day needs are so insistent that we apperceive most things merely as aids to our survival or our comfort: the rarer flashes of aesthetic vision reveal the essential characteristics of the thing as it is in and for itself, and so as possessing a value of its own. The products of human art are artistic, because "it is only when an object exists in our lives for no other purpose than to be seen" (or heard) "that we really look at it (or listen to it). It expresses something which was latent, but which we never realized." As Croce neatly puts it, "il pittore è pittore perchè vede ciò che altri sente solo o intravede ma non vede".

[2] As regards this aspect of the general problem, one of the most informative studies, I venture to believe, is still that carried out by Denman Ross of Harvard in his *Theory of Design* (1917): his final conclusion is that "a composition is of value in proportion to the number of orderly connections which it displays". The notion was subsequently taken up by G. D. Birkoff, who went on to suggest a method of assessing aesthetic value (M) on this basis, namely, by calculating "the ratio of the number of unifying elements (O) to the number of elements responsible for variety" (C): the formula thus reached is $M=O/C$ (*Aesthetic Measure*, 1933). One would have expected rather the reverse, viz., the number of elements responsible for variety *divided by* the number of unifying principles needed to organize the whole into a coherent pattern. Birkoff carried out no experimental work to verify his formula; but at my suggestion Dr. Eysenck undertook an investigation in which the 'measures' given by the formula were correlated with the preferences given by a wide sample of judges: he found that "agreement with actual expressed preferences is relatively slight". (For a fuller account, with criticisms, the reader may refer to H. J. Eysenck, *Sense and Nonsense in Psychology;* see also his Ph.D. Thesis, 1940, Univ. London Library.)

[3] There seems no ground for doubting the tradition that Pythagoras himself discovered that consonant sounds are produced by strings whose lengths stand to one another in simple numerical proportions, and that from this and various geometrical discoveries he inferred that "all things are numbers"—numbers and ratios being commonly conceived by the Greeks in geometrical rather than algebraic terms. The Greek word 'harmonia' means not 'harmony' in the modern sense, but rather the adjustment or fitting together of parts in accordance with some purpose or guiding rule; (it is a metaphor not from music but from joinery): ἁρμονία δὲ πάντως ἐξ ἐναντίων γίγνεται, ἔστι γὰρ πολυμιγέων ἕνωσις. (Philolaos, *Frag.* 10, *c.* 450 B.C.). The application of analogous proportions to the problems of art and architecture is apparently due to Vitruvius (*De Architectura*, esp. III, 1, IX, 5: *c.* 30 B.C.), who may have relied on lost Greek texts dealing with the subject. Vitruvius in a puzzling passage maintains that the proportions of the human figure ought to be followed in building temples. Interpreted in terms of the cosmology of Plato's *Timaeus* the famous 'Vitruvian figure'—a naked man set in a square and a circle (familiar from the drawings of Leonardo, Dürer, and others)—thus became a symbol of the 'mathematical sympathy' between the microcosm and the macrocosm, and Christian churches of the Renaissance were laid out on the plan of a crucified figure with out-stretched arms.

> From harmony, from heavenly harmony
> This universal frame began . . .
> The diapason closing full in man.

After nearly two centuries this rationalized method of composition grew too mechanical, and began to pall. It was, however, in England that the Renaissance conception of beauty eventually received what some have declared to be its death blow. Newton's theory of gravitation undermined the celestial scheme on which Kepler has based his *Harmonice Mundi* (1619); and the materialism and scepticism of Hobbes, Hume, and the later empiricists shattered the whole Neoplatonic cosmology. Hogarth was merely voicing the altered outlook of his day when, in his *Analysis of Beauty*, he satirized " the strange notion that, because certain uniform divisions upon one string produce harmony to the ear, similar distances in form would in like manner delight the eye ".[2] And yet the basic elements in the ancient doctrine have since shown astonishing signs of life; and, as Wittkower points out, it has once again begun to engage the attention of contemporary architects, and has exercised a manifest influence on the devotees of cubism and abstract painting.[3] We may readily agree that Alberti's formula cannot provide, as he and his followers supposed, a rational definition of beauty, nor even a sufficient condition for its emergence. Nevertheless I believe that, in the broadest possible sense, latent organization or *order* is at least a *necessary* condition. Now in its most obvious and elementary forms order can always be expressed mathematically, at any rate to a first approximation; and then even the simplest and most inexperienced mind may be able to grasp its general nature. To the more sophisticated spectator, however, a pattern that can be summed up in a mathematical equation will seem machine-made and lifeless, and so, as Hume expresses it, 'insipid and conventional'. But, whether obvious or elusive, the order that underlies all forms of beauty is not something the spectator invents, but something he finds already there awaiting discovery, and, in that sense, 'objective'.

---

[1] *De Re Aedificatoria*, 1485, Bk. VI, c. ii. Alberti's theory had a profound influence on Bramante, Michael Angelo, and above all on a young sculptor named Andrea di Pietro, nicknamed ' Palladio ' (the name of the angel in Trissino's poem who explains the divine significance of geometrical proportions); and thus we arrive at the socalled Palladian style.

[2] *Analysis of Beauty*, 1753, pp. 76, 131). Hume a few years later maintained, as we have seen, that " beauty is not a quality in things themselves . . . To reduce every expression to geometrical truth and exactness would be most contrary to the laws of criticism, because it produces work which has by universal experience been found the most insipid " (*Of the Standard of Taste*, 1757). Similarly Burke, writing in the same year, denied at considerable length that beauty " has anything to do with calculation and geometry ", and made sesquipedalian fun of the Vitruvian figure and the " forced analogy between the proportions of the human body and the proper proportions of a building " (*Inquiry*, Pt. III, sect. ii, iii, and iv).

[3] *Architectural Principles in the Age of Humanism* (1952). Mr. Whittaker is mistaken both when he quotes me as attributing Fechner's theory of the golden section to " Roman and Renaissance architects ", and when he assures us that " no present-day artist or architect would take seriously these exploded speculations of the founder of experimental aesthetics so-called ". It is true that, tucked away in a footnote to my paper on Fechner, there was a reference to the doctrines of both Vitruvius and Palladio (this *Journal*, XIII, p. 7). But the use of the golden section was not ascribed to either. On the contrary, in accordance with the Pythagorean theory the proportions which they adopted were ratios of simple whole numbers: the only incommensurable ratio they allowed was the Pythagorean ratio of the side of a square to its diagonal. Just at the time when (as Mr. Whittaker declares) " the experimental studies of 20th century aesthetic psychologists were rebutting the alleged use of the Fechnerian golden section in art ", Le Corbusier was starting to base his own designs on that very principle. His celebrated ' modulor ' is a " measuring tool derived from the human body and from mathematics: a man with arm upraised provides . . . three intervals which give rise to a series of golden sections called the Fibonacci series " (Le Corbusier, The Modulor: *A Harmonious Measure to Human Scale applicable to Architecture*, 1954, p. 55: the Fibonacci series are the numbers 1, 2, 3, 5, 8, 13, etc., where each is the sum of the last two, and the ratios of each pair progressively approximate to that of the golden section; Fibonacci, i.e., Leonardo of Pisa, is best known as the mathematician chiefly responsible for the introduction of the Hindu-Arabic notation: his

The Objectivity of Value. This brings us to the main complaint that my critics have urged against the views and assumptions that I have from time to time defended, namely (to put it in a single phrase) that I *objectify the subjective*. But in the first place it must by now be clear that much of the plausibility of the subjectivists' argument hinges on the ambiguity of their key term—'objective' (cf. p. 65). In the premiss of their argument the expression 'not objective' appears to be merely a synonym for 'subjective'; in the conclusion it is used to carry the further implication of 'unreal, and therefore non-existent'. But plainly, just because 'value' is 'not objective' in the former sense, we cannot infer that it is 'not objective' in the latter sense without committing a flagrant *non-sequitur*.

Take, by way of an example, the case of a 'negative value' such as *pain*. Following the same argumentative steps that are adopted by our critics (pp. 48f. above), we could start by pointing out that the experience of pain is "dependent on certain physico-chemical changes" in the sufferer's "body and nerves": it is therefore "subjective", and in that sense "not objective". If we then accept Mr. Whittaker's mode of reasoning, we should infer from this that pain is "wholly reducible to terms of certain physico-chemical processes". And since it is part of our basic assumption that "our ultimate conclusions must be expressed in physico-chemical terms", it would seem that pain as a mental quality must be non-existent, and so cannot possibly form an element in "absolute reality". However, I doubt whether Mr. Whittaker himself—or indeed anyone save the most stoical of Christian Scientists—would accept such a conclusion, though it is clearly the conclusion to which his assumptions here lead. Because pain is *dependent* on physico-chemical processes, we cannot infer that it can be reduced *without remainder* to physico-chemical processes. The feeling of pain is after all one of the most irrepressible of realities. Even the most thorough-going subjectivist, when racked by toothache, must own that, so long as it lasts, the sensation of pain undoubtedly exists.

---

discussion of the series which now bears his name occurs, oddly enough, in a problem relating to the multiplication of rabbits: *Liber Abaci*, 1202, A. D., 283f. The series is discussed in textbooks on theory of numbers, usually in relation to continued fractions and recurring series).

May I say how curious it is that nearly every psychologist who has discussed the psychology of beauty ignores the art of architecture (as a glance at Valentine's recent review sufficiently shows)? That the notion of proportionality is by no means so completely extinct in other fields as Mr. Whittaker seems to suppose is demonstrated by the following passage from a contemporary art critic. "Beauty is a recognizable quality; the artist no more creates beauty than the scientist creates truth; he lifts a corner of the veil and *reveals* beauty . . . It might be described as law-abiding behaviour. But he has no need to *formulate* the law. Since it exists, it must manifest itself in some kind of pattern; and in the ultimate analysis every pattern can be expressed in terms of mathematics . . . An orderly, law-abiding mathematical world is delightful in its own right." (Eric Newton, *The Meaning of Beauty*, 1959; Penguin edition, 1962, pp. 31f. and figure 3, p. 44). Let me add that, were I a Manichee, I should be tempted to base my dualism, not on the old antithesis between matter and spirit, evil and good, error and truth, but on the wider contrast between disorder (in the sense of blind chance or randomness) and order (of which the type is rational, purposeful planning).

For those philosophers who have tried to see and understand 'reality as a whole' no problem has proved so baffling as the presence of evil as an actual ingredient in the universe; and no form of evil is so insistent and intrusive as pain. How eagerly they would jump at Mr. Whittaker's solution, were it only tenable. And yet, as each of us knows to his cost, merely to declare that pain is purely subjective does not, like some simple anodyne, abolish its reality.

But if so, an argument that holds of negative forms of value must surely hold equally of positive forms. Whether happiness and pleasure, beauty, truth and goodness are or are not 'subjective', they too are unquestionably ingredients in the universe as we know it: they, too, are irreducible. And, though neither material nor physico-chemical, they are nevertheless, in one sense or another, unquestionably real. "Sunset and sunrise *are* beautiful if one child thinks them so."[1]

At this point my critics will interpose with the protest that I have now transported the argument into the realm of metaphysics; and "psychology as a natural science can have no concern with the metaphysical". To maintain that pain and pleasure, goodness and evil exist or are real, is to formulate an ontological assumption; and that, in the view of every behaviourist, is sufficient to condemn it out of hand, because, as we are assured again and again, "ontological assumptions are unamenable to any operational test and are therefore meaningless". Yet plainly their own assertion that value is not objective and therefore unreal is itself an 'ontological assumption', and certainly could never possibly be demonstrated by any 'operational test'.

The truth is, the old nineteenth-century dogma that "philosophy and science" (in the sense of natural science) "are to be kept sharply apart", and that there is "an unbridgeable antithesis between value and fact", has long ago been exploded.[2] Since the days of Einstein and Planck, modern physics has found itself increasingly obliged to discuss problems that were once held to be purely metaphysical; and modern philosophers have come more and more to base their doctrines on the findings of natural science. Psychology more than any other branch of science, is constantly compelled to take sides on philosophical issues, and to adopt, provisionally at any rate, certain presuppositions that are admittedly 'ontological'. Its ultimate aim, as I have already argued, is to explain and understand the phenomena with which it deals, in other words, to organize them and the generalizations which they suggest into an intelligible and articulate system. For this purpose it is forced from time to time to introduce tentative hypotheses about the nature of what exists or what is real—assumptions for which there can be no direct empirical check, but which nevertheless seem indispensable if we are to construct a satisfactory working explanation. Dr. Taylor and others complain that my own assumptions are often no more than 'wish fulfilments'. In a sense that is true, for we all 'wish' to find some satisfying interpretation to put on the facts we observe—some comprehensible scheme such as will guide us in solving our practical and our intellectual problems. To that extent I am unashamedly a pragmatist.

We used to be told, as Mr. Whittaker reminds us, that aesthetics, ethics, and logic are "normative sciences not positive or natural sciences: they are prescriptive rather than descriptive". That may hold good of those disciplines in and for themselves. But there is also a *psychology* of aesthetics and a *psychology* of ethics, just as there is a *psychology* of rational thinking; and it is the business of the psychologist to study the nature and origin of these so-called 'normative' principles and processes, and to reduce the observable data to an ordered and intelligible system which will provide a reasonable basis for the rules prescribed. The conclusions and criteria reached in this way may incidentally help us to criticize or refine our aesthetic and moral judgements in unusually complex or controversial cases. Nevertheless the ulterior purpose of such studies is neither to describe nor to prescribe, but to explain.

---

[1] Quoted by Sherrington, *Man on his Nature*, p. 393.

[2] Here Professor Skinner and I are in agreement, though for very different reasons; (see B. F. Skinner, *Science and Human Behaviour*, 1956, section on 'Value Judgements', pp. 428f.).

Now a purely ' behavioural ' treatment of value, such as Dr. Gregson demands, would leave the facts that have emerged from first hand empirical studies (such as the experiments reported above) utterly incomprehensible. Merely to dismiss them as " the subjects' verbal reports " is tantamount to dismissing them altogether. On the other hand, if we accept them at their face value, and adopt the interpretation they naturally suggest—namely, that beauty and goodness are independent and irreducible characteristics of the objects to which they belong, directly apprehended as such, and in that sense ' objective '—then we can make much better sense out of human life and our own very fragmentary experiences. Those who " deny that value actually exists ", seem to me to take far too narrow a view of both ' value ' and ' existence '. They think of ' value ' solely in abstract, whereas the values we directly apprehend are always *concrete* values, characterizing *individual* objects. And they identify ' existence ' with the mere specification of position in time and space (as Dr. Taylor also does), whereas the existent cannot possibly be treated as *no more* than that which can be dated and located. Indeed I myself should hold, as Whitehead has so convincingly argued, that individuality is " precisely the feature in things and in persons which resists any attempt to state its spatial and temporal coordinates ".

Thus the behaviouristic theory does not merely fail to explain the facts of ordinary life; it falsifies them. In ordinary life actuality and value always appear as aspects which, though distinguishable in theory, cannot be severed in fact. And it is the function of philosophy to demonstrate their interrelation. " Die Metaphysik macht es sich zur Aufgabe die gesamte Welt, in der es Wirklichkeit und Werthaftes giebt, einheitlich zu deuten."[1] The beauty which has inspired so many painters and poets, the ideals of goodness which have moved so many heroes to superhuman efforts—these have exercised their power, not because they are mere subjective illusions imposed on a set of objects or events that really consist of nothing but quantified particles and their conjoint motions, but because beauty and goodness are essential constituents of reality itself. Even were the universe itself purposeless and blindly indifferent to either beauty or morality, it is nevertheless a universe in which morality can arise and flourish, and in which beauty can be recognized, created, and enjoyed. " What confronts us in actual life ", says Professor Taylor, " is neither facts without value, nor values attached to no facts, but fact revealing value.[2] Mr. Whittaker may call such a view " just a naïve and uncritical impression " if he likes; yet unless we adopt it or something like it as our working hypothesis, we make nonsense not only of the universe but of all human psychology and history.

---

[1] H. Rickert, *System der Philosophie*, 1892, p. 138.

[2] A. E. Taylor, *The Faith of a Moralist*, p. 61.

As a mere psychologist I am scarcely competent to answer the frequent denunciations of ' theological categories ' in which Dr. McLeish indulges in his recent book (*The Science of Behaviour*, pp. 1, 191f.) and which are so vigorously hammered home by Mr. Whittaker in his letter above. But I may perhaps be allowed to make two passing observations. The notion that we directly perceive value as an objective quality of things and of actions survives in psychology (we are told) merely as a relic of the traditional theological doctrines imported into it by Christianity. But here, I fancy, those rationalists or humanists who are not already committed to Watson's mistaken notions of historical theology would say exactly the opposite. Santayana, for example, maintains that theology and particularly Christian theology " accustomed people to disregard the intrinsic value of things ". As symbols of the heavenly verities things could have a borrowed value only, and good actions had value only so far as they were instrumental in procuring salvation. Intrinsic value was to be found not in earthly existence but solely in a spiritual kingdom in the sky (*Realms of Being*, IV, 1940, pp. 208f. and *Little Essays*, ed. L. P. Smith, 1920, ' Intrinsic Values) '. Secondly, so far from supporting Mr. Whittaker's contention that " the existence of value is a doctrine or dogma deduced from, and dependent on, the traditional belief in God ", the only plausible argument connecting the two must surely proceed in the opposite direction. If any inference is permissible it is rather that the hypothesis of a moral deity ' follows from the fact that value is an essential element in the universe; and if, as he says, Reality, it appears, requires a binocular vision to see it as it is—in its depth as well as in its breadth. It needs, so to speak, a stereoscopic combination of the *factual* view of the scientist and the *evaluative* view of the poet, the artist, and the moralist. A one-eyed survey which recognizes either nothing but facts or nothing but values is bound to be grossly deceptive. " The principles of operational research ", we are told, " possess an almost uncanny power to reveal the deepest secrets of Nature and to add to the wealth and welfare of mankind ". That may be true so far as material phenomena are concerned. But the attempt to apply the same techniques to the phenomena of value, and so reduce them to " subjective characteristics of material things ", will in the end deprive both of all meaning.

There is an allegory told in *The Arabian Nights* by a wandering faqir which describes how one of the pilgrims to Mecca purchased a box of magic ointment. When he anointed his right eye with the salve, he could see the most precious treasures hidden in the remotest corners of the Caliph's realm. But when he anointed both his eyes, he was suddenly struck blind.

V. SUMMARY AND CONCLUSIONS

May I now try to summarize the outcome of this long and intricate discussion?

1. In the past numerous attempts have been made by psychologists and philosophers to reduce what are commonly known as ' values ' to terms that are essentially non-valuative —to identify ' beauty ', for example, with what we like or what gives us pleasure, ' goodness ' with the greatest happiness of the greatest number or with the behaviour enjoined by social traditions, or both with what arouses an emotion of approval, and so on. Accordingly in the foregoing pages I have attempted to classify the various modes of reduction, adopting a strictly psychological standpoint, and then briefly examined the arguments for and against each one. The general upshot is that all such proposals crumble and collapse in the face of what seem to be insuperable objections. The more recent efforts by contemporary psychologists—particularly those of the psychoanalytic and behaviouristic schools—to effect a similar reduction, and so to render the concept of value superfluous, turn out to be little more than revivals of methods suggested in the past (partly disguised by changes in terminology), and are usually put forward regardless of, and often, it would seem, in ignorance of, the numerous criticisms with which such views have been riddled by later writers. Since therefore all attempts at reduction have hopelessly failed, we may fairly conclude that the concept of ' value ' is unique and irreducible.

2. From this it would seem to follow that ' value ' and its different modes or forms— ' goodness ', ' beauty ', ' truth ' and the like—are logically primitive, and therefore indefinable. In that respect they resemble other relatively simple qualities that are directly experienced, such as pleasure and pain, yellow or pink, the scent of a rose or the flavour of a vintage port. Nevertheless, though we cannot define them, we can indicate what we mean by pointing to concrete instances in which such qualities are conspicuously present or conspicuously absent, or in which they are exhibited in conspicuously different degrees. The only clue that has emerged is the observation that all three types of intrinsic value involve some kind of *order*. This means that the appreciation of value implies a level of ability that is capable of apprehending relations.

3. To gain further light on the issues thus raised I have endeavoured to carry out a series of experimental studies, based on a wide variety of materials and a representative

" value in any case could only be of value in or of a conscious person ", then that would, if anything, seem to imply that " the architect of the universe " was at least a conscious person (though presumably something very much more). I hasten to add that this is not an argument I myself should put forward, because I do not accept the premiss. Those who are interested in this aspect may be referred to the detailed discussion in W. R. Sorley's Gifford Lectures on *Moral Values and the Idea of God.*

panel of subjects from every walk of life, using not only quantitative methods such as paired comparison and ranking, but also an informal questionnaire to interpret the data. The empirical evidence so obtained fully confirms the view that value, and differences in value, are things that people directly apprehend; but the mode of apprehension seems to resemble that of Gestalt-like qualities such as shape or order, rather than that of so-called secondary qualities such as colour, sound, or smell. Their judgements about beauty, and still more their judgements about goodness, they all, with few exceptions, regard as statements of fact—assertions whose truth they would defend against any who ventured to contradict them.

Beauty thus appears to be a quality which we discover in certain objects and goodness a quality which we discern in certain actions; they are not unauthorized additions that we somehow impose on, or project into, the objects or the actions, as a result of our own private feelings or desires. It follows that the experience of value is primarily a process of a cognitive type, though, like all cognitive processes, it may be accompanied by affective or emotional processes, or be influenced by conative or appetitive tendencies, or modified by personal associations.

4. As a result of these subsidiary influences, there are often marked individual differences in men's judgements about beauty or goodness, just as there are in their judgements about other perceptible qualities or about the truth and validity of a logical deduction. Nevertheless, the correlations between the rank-orders obtained from different individuals are almost invariably positive, and much larger than is usually assumed. With a sufficiently wide range of objects or instances, the amount of general agreement completely outweighs the amount of disagreement.

5. Could we accept the foregoing results as they stand, the natural inference would be that value, like shape and differences in shape, exists independently of human recognition, and in that sense is genuinely objective, or at any rate more objective than subjective. The evidence is admittedly by no means conclusive, if only because the issue is partly verbal as well as observational. Yet it furnishes a strong prima facie case; and the inference implied is in keeping with the almost unanimous verdict of common sense. The burden of rebutting it is thus shifted onto the backs of those who want to discard it. Merely to keep awkward questions at bay by flourishing Occam's razor is no adequate answer; and to deny the existence of values on the ground that this could involve a metaphysical assumption is itself a metaphysical assumption. The duty of any would-be critic therefore is to confront us with empirical evidence on the opposite side; and this he has so far failed to do.

Meanwhile, this broader standpoint at once suggests a wide variety of unsolved problems for scientific research which current behaviourism merely ignores or waves aside. If we adopt these natural inferences as a working hypothesis, it becomes far easier to interpret, not only the more elusive facts of aesthetic and ethical experience, but also the many basic issues of policy and conduct which occasion so much doubt and disagreement. The so-called behaviourist hypothesis, when it does not just ignore such phenomena, contends that they can be analysed without remainder in terms of simple discriminative or emotive responses; but all such attempts still leave the fundamental facts entirely unexplained. Finally, the objective view that I have defended brings with it practical advantages as well as theoretical. It provides a far more effective basis for the work of the practising psychologist in the various fields in which such issues are especially apt to arise—in dealing with delinquents or neurotics, in handling social and industrial relations, and in unravelling the moral conflicts that challenge every one of us in the course of our everyday life. The subjective view leaves the psychologist in the position of the mariner who tries to navigate his ship by a landmark tied to his own mast head.

But now, in the words of one of Jules Verne's heroes on a memorable occasion, "Enough of speculative pros and cons; what we need are facts—more facts and relevant facts; let us therefore plan some first-hand experiments, and hope that the results will bring us nearer to a conclusion which all can accept".

TABLE I. EXPERIENCES OF VALUE

The figures are percentages to nearest whole number. For further explanation see pp. 85–90. (The arabic numerals attached to side-headings correspond with numbers in text.)

	O-group	S-group
A. *Physical*		
1. *Comparison of Size*		
(i) Differences	100	100
(ii) How known		
Direct perception	84	71
Inference	3	16
(iii) Differences independent of recognition?	95	91
(iv) Quality	97	96
(v) Quality independent of recognition	98	89
B. *Aesthetic*		
2. *Comparison of Pictures*		
(i) Differences	95	97
(ii) How known		
Direct perception	90	81
Inference	0	4
Emotional or affective	3	6
(iii) Differences independent of recognition?	96	73
(iv) Beauty independent of recognition		
Yes	81	72
No	2	11
Doubtful	10	14
3. *Natural Objects* (*Flowers*)		
(i) Beauty: how known		
Direct perception	77	65
Reflection, inference, etc.	0	6
Emotional or affective	16	28
(ii) Shape independent of perception	94	87
(iii) Colour independent of perception	86	63
(iv) Beauty independent of perception		
Yes	82	48
No	0	14
Doubtful	5	39
4. *Comparison of Natural Objects* (*Landscapes*)		
(i) Differences: how known		
Direct perception	73	59
Inference	2	8
Emotional or affective	15	23
(ii) Differences independent of recognition	91	45

	O-group	S-group
C. *Ethical*		
5. *Kindness* (*The Good Samaritan*)		
(i) Grounds for judgement		
Direct recognition	76	53
Rational grounds	21	31
Emotional grounds	0	12
(ii) Judgement of fact	97	88
(iii) Differences independent of recognition	81	72
6. *Courage*		
(i) Grounds for judgement		
Direct recognition	68	41
Rational grounds	7	32
Emotional grounds	12	17
(ii) Judgement of fact	76	69
(iii) Differences independent of recognition	70	57
7. *Cruelty*		
(i) Grounds for judgement		
Direct recognition	97	86
Rational grounds	3	5
Emotional grounds	0	9
(ii) Judgement of fact	98	93
(iii) Fact independent of recognition	98	93
D. *Experiences of the Sublime*		
(i) Frequency	18	47
(ii) Occasions		
Ordinary natural phenomena	8	16
Exceptional natural phenomena	3	35
Personal situations	15	12
Death or danger	7	10
Love	5	4
Ostensibly extra-sensory	2	8
Human achievement	4	6
Literary	2	23
Musical	5	18
(iii) Characteristics		
Sense of external power	14	47
Sense of personal impotence	16	31
Feelings of fear, depression	8	4
Feelings of joy, exaltation	10	43
Sense of external purpose	9	18
Sense of personal detachment	12	36
Sense of union	7	11

ABILITIES AND ATTAINMENT. I

To the Editor, The British Journal of Statistical Psychology

SIR,—Dr. Levy's paper in your last issue starts from the note by Pidgeon and Yates published in the N.F.E.R. bulletin No. 8 (1956). Since, as the authors record, this note in turn stemmed from some remarks I made in conversation with them, and since these remarks are relevant to Dr. Levy's paper, perhaps I may be allowed to recall what I actually said. I began by quoting a well-known passage:

" Intelligence, as psychologists use the word, is almost synonymous with what the Board of Education, somewhat uneuphoniously, has called ' educable capacity '. Capacity must obviously limit content. It is impossible for a pint jug to hold more than a pint of milk; and it is equally impossible for a child's educational attainments to rise higher than his educable capacity permits. In theory, therefore, we should always expect his educational age to fall below his mental age. In practice, however, where standardized tests are applied, the former is found to exceed the latter in a few rare instances ": (Sir Cyril Burt, *The Backward Child*, p. 447).

This passage, I went on to say, could hardly support the meaning commonly given to it, since that would imply a geometrical impossibility, with any of the scoring systems in actual use: it would imply (I went on) with any of these systems, that it is possible to draw a straight line through the centroid of a cigar-shaped cloud of points in a plane such that nearly all the points are on one side of it. This is patently absurd. It is only possible for a line through the centroid of a distribution to divide the points into a few on one side and all the rest on the other if the distribution is like a comet, with a short heavy head on one side and a long light tail on the other, which is not remotely like the bivariate distributions of test scores given by the systems of scoring that are actually used. Dr. Levy's remark that " if we equalize the means and standard deviations of two distributions where the correlation is less than perfect, then a considerable proportion of the differences would be greater than zero, and an analogous result would obtain for the ratios " is a special case of this more general proposition.

The favourite remark of one of the best teachers of mathematics that I have known in the course of thirty years as H.M.I. was " Let's have a few for instances ". It was with this remark in mind that I went on to urge the authors to include one or two examples of actual bivariate distributions in their note. Although, I said, these examples added no logical cogency to a self-evident geometrical fact, they would none the less be found surprising by many readers. This forecast was justified by the event, as may be seen from the subsequent correspondence.

If the scoring system is modified, so as to remove the restriction that the line separating over achievers from under achievers must pass through the centroid, it is no longer the case that the number of under achievers must be roughly equal to the number of over achievers. We can add a hundred points to each I.Q., and everybody becomes an under achiever. Or to each A.Q., and everyone is an over achiever.

' He wished all men as rich as he
And he was rich as rich could be
So to the top of every tree
Promoted everybody.'

Dr. Levy proposes the former convention, or something rather like it, when he speaks of 13-year-olds being on the average at least one year retarded. This convention has the advantage of drawing our attention to the fact that our educational world is not as good as we may hope that it will become. The opposite convention would draw our attention to the fact that it is now better than it once was. The centroid, or " equal means ", convention that is actually in use is equally arbitrary, and would be equally harmless, but for the misconceptions that have clustered round it. In short, the choice of convention as to means is arbitrary, as Dr. Levy says explicitly later in his paper. This is clear enough. What is less clear is the reason why he wishes to "postulate the properties of normality and equal variances within the arrays". Postulation is an uncomfortable word in this context, reminding one of Russell's remark that the advantages of postulation over proof are those of theft over honest toil; and in any case it seems unnecessary in view of the rest of the paragraph.

While I agree with Dr. Levy's views about the arbitrary nature of the scoring systems I find it hard to follow his remark that "measures of ability should be chosen so that the psychological functions they tap are, relatively speaking, more stable than scholastic achievement: this is a sufficient requirement". Sufficient for what purpose? And do the measures in actual use have this property? It was doubts on these points that caused me to raise the whole question with Pidgeon and Yates in the first instance, and during the intervening eight years I have seen no evidence to remove these doubts. Dr. Levy's final paragraph indeed inclines me to think that he shares them, for, as he says, "there remains the problem of showing that the underlying concepts of the theory and its implied numerical operations are meaningful in practice". Or, to put the matter another way, can this psychometric model be used effectively to supplement the clinical judgement of skilled psychologists, and so enable their work to be more fruitful? It must be remembered that the value of any new information is always dependent upon what is already known, so that what might be useful information in some circumstances ceases to be so in others. Whether the model yields information that is both new and useful can only be found out by actual experiment, and cannot be inferred from the design of the model. If the tests of ability and attainment have the properties that we hope they have, then the experiments will be successful. If not, not. Experiment has shown that these tests have many useful properties; but it has not yet shown, or so it seems to me, that they have those that are in question here. This was what I had in mind when I originally raised the matter; and, so far as I am aware, the question is still unanswered.

G. F. PEAKER.

ABILITY AND ATTAINMENT. II

I fully agree with Mr. Peaker that the passage which he and others have quoted from my book on *The Backward Child* "could hardly support the meaning given to it" by several of my critics, including Mr. Brimer in one or two of his earlier articles. The book, published more than a quarter of a century ago, was concerned almost exclusively with the educationally subnormal, and was intended for the practical teacher and for students in training. It was therefore written (so to speak) at 'O-level' rather than 'A-level'. The more abstract theoretical problems that interest the statistical psychologist had already been discussed, so far as they seemed relevant, in papers previously published in technical journals or in preliminary memoranda. Mr. Peaker's quotation consists of the opening words of a section (in the second half of the book, p. 447) headed 'Educational Effects of Lack of Intelligence'; and, as the footnote indicates, it was a recapitulation of an argument and a problem set forth at greater length near the beginning of the volume (pp. 34f.), where the terms employed were more fully explained. The continuation of the extract leads up to the conclusion, stated at the end of the section, that, although the most important single cause of educational backwardness appeared to be an innate lack of general intellectual capacity (a milder form of the deficiency that produces what in those days was certifiable 'mental deficiency'), nevertheless deviations in 'general intelligence', least of all deviations as assessed by fallible tests, *formed by no means the only influence* affecting children's educational progress, particularly when this too is estimated by simple scholastic tests or by the traditional type of scholastic examination.

The theoretical investigator, whether he is using experimental techniques (such as mental tests) or statistical procedures (such as factor analysis) or a combination of both, is apt to oversimplify the various issues involved. In dealing with individual children (as I myself, for example, had to do in my work for the London County Council) it is never sufficient to apply a single test of intelligence and use the scores just as they stand. The test-results, particularly in the case of the subnormal, must be checked against teachers' individual estimates, and in some cases corrected by the application of an alternative scale (performance tests, or the like). Besides intelligence, other abilities or disabilities must also be assessed. Motivation and various character qualities must be taken into account; home conditions ascertained; and, if possible, the abilities of relatives (not parents only) must be estimated. It is the neglect of these supplementary observations, a neglect to which statistical investigators seem particularly prone, that tends to bring the procedure into discredit, and so leads to the sweeping condemnations pronounced by onlookers—like Dr. McLeish, for example—who dismiss the whole race of mental testers as blind leaders of the blind, " out of touch with reality and failing to say anything significant about human behaviour " (*Science and Behaviour*, 1963, pp. 29f.).

At the time when *The Backward Child* was written, the prevalent theory, expounded in training colleges by lecturers on educational psychology, was that of Professor Spearman, usually without the incidental qualifications he himself introduced. This theory placed almost exclusive emphasis on a ' general factor ' of intelligence, and dismissed the notion of special abilities (or ' group factors ') as " an obsolete survival of the old-fashioned faculty psychology ". Such an interpretation, largely as a result of its attractive simplicity, gained for a while enthusiastic adherents. At the same time the broad popular meaning given to such terms as ' intelligence ' and ' capacity ' engendered further misunderstandings. It was for this reason that, in the chapters following the one quoted by Mr. Peaker, I laid such stress on ' special abilities ' and ' temperamental and emotional characteristics '.[1]

The issue to which Mr. Peaker refers came first to the fore as a result of an investigation which I carried out in 1919 with the help of a number of teachers trained in research methods. The results, published in a report on ' The Relations between Mental Ability and Educational Attainments ', were subsequently incorporated in *Mental and Scholastic Tests* (1921, sect. 6, pp. 175–180 of 1st ed., pp. 219–225 of 4th ed., 1962). Out of the 689 children covered by the survey there were only 63 whose educational ratios were more than 10 points above their mental ratio, and as many as 166 whose educational ratios were more than 10 points below it. The latter figure occasioned no comment; but the former excited considerable surprise and discussion. If a child's intelligence quotient measures his ' educational capacity ', how is it possible for his educational attainments to *exceed* his capacity. Is not this (so I was asked) " the conjuror's trick of pouring more than a quart of wine out of a quart bottle "? (cf. *The Backward Child*, p. 35 and refs.).

The answer is quite simple. Educational attainments, as we have seen, are affected by many other factors besides ' general intelligence '. Consider children whose I.Q. is exactly 100. We measure their educational ratio or E.Q. An E.Q. of 100 would then indicate the attainments of a child whose special abilities and other characteristics (as well as his I.Q.) were all approximately average. If, however, a child whose 'intelligence' is merely average happens to be endowed with a specific 'numerical ability' and a degree of retentiveness well *above* the average, and if in addition his motivation is exceptionally good and he receives individual help and encouragement from his teacher (as in several of the illustrative cases described at length in my report), then his performance in the arithmetic test will be well *above* the average. Similarly with other educational tests. In such a case therefore—paradoxical as it may seem to those who overlook these supplementary influences—the child's E.Q. will *exceed* his I.Q. If, on the other hand, his memory, his special ability, etc., are all below the general average, then, even with an I.Q. of 100, his E.Q. will fall well below that figure.

---

[1] Dr. McLeish complains that ' intelligence testers ' like myself commit three methodological errors which " vitiate this work ": (i) we ignore ' perception ' in favour of abstract processes such as reasoning; (ii) we ignore the influence of ' will ' (i.e., motivation) and of ' emotion '; and (iii) we seek to estimate ' potentiality ' instead of assessing ' the *actual* intelligence of the child in his present environment ' (*loc. cit.*, pp. 39f., his italics). I seems clear, however, that he has overlooked both the chapter headed ' Perceptual Processes ' which immediately follows the chapter on ' Intelligence ', and the final chapter in the series headed ' Temperament and Character ', which emphasizes and illustrates the influence of emotional and motivational factors. In dealing with the backward child it would be plainly insufficient to stop at assessing ' *actual* intelligence in the present environment ' (which, of course, a psychologist would not call ' intelligence ', since that term is reserved for a technical concept). Deficiency in ' *actual* intelligence ' merely indicates a problem; and the psychologist's first requirement is to discover how far this deficiency is remediable and how far it is due to an irremediable defect in genetic constitution, i.e., in what Dr. McLeish terms ' potentiality '.

The most striking cases in which children's attainments exceed what one might expect from their scores for intelligence are to be found, not in schools under the local authorities, but in independent preparatory and public schools. As I noted when reporting test results obtained at Rugby and other public schools, the range of ability extends to a much lower point than is commonly imagined: at 110 I.Q. the proportion drops rapidly to about 10 to 20 per cent, but boys are occasionally found whose I.Qs. are as low as 105; and many of these less intelligent pupils do far better academically than might be anticipated. This seems confirmed by the well known fact that boys in public schools, who have failed at the local 11 plus examination, may not only obtain two or more passes at 'A-level', but occasionally secure scholarships at a university. For the most part they profit, not only by their excellent home backgrounds and the attention given in small classes by highly skilled masters, but also by their own personal qualities—memory, industry, ambition, and the like.[1] It seems clear therefore that, if we cared to make a more searching study of the individual capabilities of children who were just below the present 11 plus borderline, and if we could bestow on them all the advantages enjoyed by the children at these independent boarding schools (compensating where necessary for unfavourable home conditions), we should find a far larger number in the general population who were capable of profiting in some degree by a grammar school education. But the practical question is—would the resulting gain be commensurate with the cost of providing all these extra facilities?

Nowadays, however, what puzzles our critics is apparently not the occurrence of 'over-achievers' but their paucity. How is it that comparatively few have E.Q.s which exceed the average and so many have E.Q.s which fall below? The answer is that the distributions are not symmetrical; they approximate to the skew distribution known as Type IV with an elongated lower tail. And the reason for this is that factors tending to reduce the child's educational performance in all his subjects are far commoner and far more influential than those tending to raise it. If, for example, a child has exceptionally *good* health, that of itself is not likely to improve his attainments; if, however, he has exceptionally *poor* health, this will lower his attainments in all school subjects. The same is true, though in a lesser degree, of school attendance, home background, industry, and the like. Indeed, in arithmetic even an exceptionally high degree of special numerical ability is not likely to make a very great difference, since that alone will rarely ensure promotion to a higher class where he can learn how to do the harder sums.

---

[1] Too much stress should not be placed on the rather large number of children from independent schools who fail in the local 11-plus. (i) To begin with, since their future does not depend on their prowess in this particular examination, they lack the strong motivation of pupils at primary schools. (ii) They seldom receive the practice and preparation for the 11-plus test that is nowadays given to most candidates at primary schools. (iii) The content of the tests—not only for intelligence, but also for English, and particularly for Arithmetic—is adapted to the curricula of the primary schools and to the cultural background of their pupils, and not to pupils at independent schools. (iv) The tests set by many local authorities are often compiled by members of the staff who are not trained in the construction of tests, and hence are far from reliable. (v) There is always the possibility that the bright child, whatever school he comes from, may fail to do himself justice at the 11-plus examination owing to temporary ill-health or emotional disturbance. Nevertheless, quite apart from these 11-plus failures, there is plenty of evidence that pupils from the independent schools, whose intelligence is undoubtedly below the usual borderline for entrance to a grammar school (about 112 I.Q.), are able, in virtue of their special advantages, (particularly, I suggest, that command of the English language which is acquired in a cultured home) to make reasonably good progress at work that is roughly of a grammar school level.

I may add that in a recent study of 'Achieving and Under-achieving High School Boys of High Intellectual Ability' (*J. Educ. Research*, LIII, 1960, pp. 172–180) much the same supplementary causes were noted as were reported in similar studies carried out in L.C.C. schools.

Mr. Peaker goes on " to urge authors to include one or two examples of actual bivariate distributions in their notes ". For the survey I have mentioned such a distribution, giving the frequencies in detail, was in fact published in the volume already cited (Table XIX, p. 177). That, however, may now seem a little out of date, and in any case has apparently escaped the notice of many of my critics. Accordingly I print below a set of figures obtained some 15 years later by much the same methods for a group nearly twice as large—1,253 in all. In the earlier inquiry the mental and educational ratios or ' quotients ' were computed direct from the mental, educational, and chronological ages of the children concerned. Thus calculated the distribution of the mental ratios is by no means perfectly symmetrical, and the distribution of the educational ratios is markedly skewed: no child had an educational ratio over 130, but five had educational ratios less than 70. This of itself prevents the bivariate distribution from conforming to the conditions needed for a normal correlation surface. Since, however, Mr. Peaker has inquired more specifically about the type of distribution found in such cases, it will simplify the discussion if we rescale the scores for both intelligence and attainments so as to remove these particular sources of distortion. Accordingly for each of the two variables a strictly normal distribution with a standard deviation of 15 points has been assumed, and every assessment has been based on the child's percentile rank. The term ' quotient ' (or ' ratio ') thus becomes a purely conventional label.[1] At the same time, for the sake of simplicity, I have reduced the frequencies themselves to proportions *per mille*, omitting decimals (see Table I.)

TABLE I. COMPARISON OF MENTAL AND EDUCATIONAL RATIOS

Mental Ratio	Educational Ratio														Total
(I Q)	<70	70–	75–	80–	85–	90–	95–	100–	105–	110–	115–	120–	125–	130–	
70	11	5	1												17
70–	3	14	17	1											35
75–	2	13	12	19	1										47
80–	1	1	10	25	31	3	1								72
85–		1	2	17	35	39	3								97
90–		1	2	3	14	46	41	4	1						112
95–			1	2	10	12	51	40	3	1					120
100–			1	3	3	5	13	45	44	4	2				120
105–				1	2	4	6	17	38	42	2				112
110–			1		1	2	2	9	19	31	29	3			97
115–				1		1	2	4	5	12	25	21	1		72
120–							1	1	2	6	8	13	15	1	47
125–										1	5	9	18	2	35
130											1	1	1	14	17
Total	17	35	47	72	97	112	120	120	112	97	72	47	35	17	1000

It is obvious at once that this later distribution is far more compact. The earlier survey was carried out at the commencement of my work for the Council, before teachers had begun to use intelligence tests themselves. And in that survey, as may be seen from the table quoted, the discrepancies sometimes amounted to 40 points or over; there were, for instance, a couple of pupils (both from extremely poor homes) whose general ability was equal to that of a grammar school candidate, and whose attainments were so retarded that they could well have been set down as dull and backward. Nothing approaching this type of scattering is discernible in the later survey.[1]

---

[1] The transformation of scales with a view to simplifying subsequent calculations is a familiar device in statistical analysis. Dr. McLeish has declared that " the fundamental basis of the intelligence test is the assumption that intellectual ability is distributed in accordance with the curve of normal probability: without this assumption it is impossible to calculate intelligence quotients " (*op. cit.*, p. 63). But plenty of people have calculated such quotients without making any assumption about the distribution, and, when they are calculated in the ordinary way, by simply dividing mental age by chronological, the quotients are *not* ' distributed in accordance with the curve of normal probability ': there is a significant divergence. The assumption of strict normality is only made as a matter of convenience for certain very limited purposes, e.g., to enable the investigator to use the simple algebraic formulae which then result or the detailed frequencies available in tables.

Nevertheless, in spite of the normalization of the two variables, the abnormality of the bivariate distribution is still perceptible. The larger contour ellipses are slightly flattened on the upper or right hand side and tend to bulge on the lower or left hand side and there is a ridge of high frequencies just above the means. This, with the increased concentration in the top left hand quadrant, certainly yields something very like the 'comet-shaped' distribution (instead of a 'cigar-shaped distribution') to which Mr. Peaker alludes. A somewhat similar distortion was in fact discernible in Table XIX; there, as I observed in my original report, it seemed to be 'wing-shaped'; the concentration at the lower ratios (top left hand corner) was not sufficient to produce anything like the 'short heavy head' to which Mr. Peaker refers. Abnormal distributions of this type are by no means uncommon. They are, for example, well marked features of the bivariate tables showing the correlation between height and weight.

From an analytic standpoint the two main characteristics of any bivariate distribution that is strictly normal are (i) that the standard deviations of parallel arrays are all equal, and (ii) that both regressions are linear. The first condition is plainly violated by the frequencies shown in Table I. The range of an array for any I.Q. above the mean is wider than that of the array for the I.Q. which falls by an equal amount below the mean, although the totals are the same, e.g., the 47 children whose I.Q.s are 20–25 per cent above the mean are spread out over 8 classes for educational attainments (95 to 130+), whereas the 47 children with I.Q.s 20–25 per cent below the mean cover only 5 classes (>70 to 90). Secondly, the mean educational ratios for most of the arrays near the middle of the I.Q. scale (85 to 105) are slightly above the value required for strict linearity, while the means for nearly all the arrays towards either end of the scale are slightly below. The discrepancies are quite small; but, owing to the high correlation and the large size of the sample, they are here fully significant. The correlation coefficient is ·917, and the two correlation ratios ·929 and ·934 respectively. Applying Fisher's criterion, we find the corresponding values for z are 1·292 and 1·505. Taking as degrees of freedom $v_1 = 12$ and $v_2 = 986$, the 0·1 per cent point for z would be about ·510. In both cases, therefore, the regressions depart significantly from strict linearity. It is these two abnormalities—the increase in the standard deviations from lower I.Q.s to higher and the curvature of the regression lines—which are responsible for the wing-shaped distribution.

From a table in which the values are tabulated by small subclasses it is not easy to deduce the numbers of children who may be reasonably termed 'over-achievers' and

---

[1] It would be tempting to cite this change as evidence of the practical value of intelligence-testing, and so to some extent as a reply to the objections raised by Dr. McLeish. However, in the present case the high correlation is doubtless due to a variety of converging factors. First of all, the method of standardizing the tests must be partly responsible. Secondly, the results of the initial test were, as usual, referred to the teachers for criticism, and thus the teachers' attention was specifically drawn to the intelligence-scores of the more exceptional children. Thirdly, it should be remembered that, as indicated above, the survey was carried out between the wars when the curricula and the teaching methods in the borough surveyed were fairly uniform. Here, however, my object is not so much to demonstrate the practical value of such tests, but merely to select an empirical table which would throw light on the particular issues raised by Mr. Peaker. Let me add that it is much to be regretted that, so far, during these post-war years no large-scale surveys of this kind have been carried out. Some results recently obtained by several of my research students from limited inquiries of their own suggest that at the present day the correlations would be much smaller. I am much indebted to Miss Baker for the calculations in the text.

‘ under-achievers ’.[1] Plainly, since the range of each class in the table is only 5 points, we can hardly apply such designations to those whose educational attainments are only one class above or below that in which they are placed for intelligence. If therefore we require the deviation to amount to at least two classes, then we might deduce from the table that the number of ‘ under-achievers ’ is 97, and the number of ‘ over-achievers ’ only 32. However, this method of determining the dividing lines is not very accurate: a difference of two classes may mean anything from 6 to 14 points. It is better to turn back to the initial data and compare the actual ratios for each child. As before, we may take a discrepancy of 10 points or more as significant; we then find that the numbers prove to be 102 and 28 respectively, that is, proportions of just over 10 per cent and rather under 3 per cent of the total sample.

With Mr. Peaker's remarks about Dr. Levy's model I am in general agreement. He seems, however, to doubt the practicability of Dr. Levy's contention that we should choose our measure of ability in such a way that it will be “ more stable than the estimates of scholastic achievement ”, and asks whether any measures in actual use do have this property. With the procedure adopted in my own work (group tests checked by teachers, and supplemented where necessary by individual tests, partly non-verbal), the ‘ reliability ’ or ‘ consistency ’ coefficients for assessments obtained with an interval of 6 to 8 years amounted to about ·80 or a little above; for the scores of group tests taken as they stand they were ·75 or a little below; for measures of educational attainments between ·55 and ·60 in the earlier surveys, and between ·60 and ·65 in the later.

Mr. Brimer's criticisms are more drastic. In the passage quoted by Dr. Levy he apparently holds that there are “ at least as many children whose attainment scores exceeded their intelligence scores as there are children who show the reverse characteristic ”. I suppose that with certain methods of testing (e.g., purely verbal group tests of low reliability unchecked by teachers) approximate equality might conceivably be encountered. But, as is shown by the tables described above, this was by no means the case in my own surveys. However, the marked discrepancy between the large number of under-achievers and the small number of over-achievers diminished appreciably as intelligence testing came into wider use among teachers.[2] My chief regret is that neither Dr. Levy nor Mr. Brimer offer us any *actual results* to demonstrate the merits of their rather different models. Of the two, Dr. Levy's (defined by the two equations marked (i) on p. 138 of his paper) plainly fits my results best, since the additional term f (which may be either positive or negative) allows for the supplementary factors that I have mentioned. Whether his further assumption that $x'=y'$ (eqn. 3) can also be accepted, must depend on the way he defines a 'true score for ability', and we are never explicitly told what precisely the term 'ability' is intended to cover. In any case, until relevant data are available to check the conflicting inferences, the issues raised by Mr. Brimer, like those mentioned by Mr. Peaker, must, as Mr. Peaker says, "remain unanswered". CYRIL BURT

---

[1] These are not the terms I myself should use. Strictly, I suppose, we should define an ‘ over-achiever ’ as one whose assessment for attainments or ‘ achievement ’ exceeds his assessment for intelligence even by the smallest quantity, and similarly for the ‘ under-achiever ’. Were the bivariate distribution normal, the dividing line would in that case be identical with the mean square *orthogonal* regression line; and this would pass through the centroid in the way described by Mr. Peaker. If, however, we are to keep to a dichotomous classification for attainments, I should prefer to take the *ordinary* regression line. This would cut each array at the mean for children of the same mental ratio (or I.Q.), not at the mean for those of the same chronological age. Accordingly, when (as here) the distribution of the arrays was skewed, there would be fewer children above the mean than below it (since each *mean* would be above the corresponding *median*).

However, a simple dichotomous classification is really of no practical interest. The reliability, even of the most accurate tests, is always less than perfect. Hence a mere repetition of the *same* test would yield ‘ over- ’ and ‘ under-achievers ’ as thus defined. *Minor* divergences between assessments for intelligence and assessments for attainments therefore can have little stability and still less practical importance. We need to distinguish both the ‘ over-achievers ’ and the ‘under-achievers ’ from the ‘ average achievers ’, and this implies taking a divergence of at least 10 or 15 points above or below the precise arithmetical mean, as I have suggested throughout this note.

[2] It would seem to be my emphasis on the relatively small number of over-achievers that chiefly aroused the incredulity of my various critics. But, as the context makes clear, in the passages quoted from *The Backward Child*, I was referring specifically to the rarity of over-achievers “ among the dull ”. Of course, if the intelligence of the dull is tested by group tests of a purely verbal type, the I.Q.s would be much lower, and something like the type of distribution expected by Mr. Brimer

ABILITY AND ATTAINMENT. III

Dr. Levy has undoubtedly performed a most useful service by expressing in the precise language of psychometrics a model for the retardation theory which my colleagues and I questioned (Brimer [1,2], Morris [4], Pidgeon and Yates [6,7]): he makes explicit a logical error which we found difficult to hold constant for scrutiny, though unhappily he seems to be one of its victims. I should like, however, in this note to point out the inconsistencies of the premises from which Dr. Levy's chain of algebraic argument stems and the consequent falsity of the equations he derives, and then to comment briefly on one or two issues that he has raised in his references to the foregoing articles by my colleagues and myself.

It is perhaps necessary to emphasize at the outset that such terms as 'true score' must have an unequivocal meaning in any psychometric argument. To that end the definition must remain consistent throughout. Thus, if α is the defining property of the class of 'true scores', and x' and y' are asserted to be members of that class, then x' and y' must be shown to possess the property α, and the property must be so defined that we can determine whether or not x' and y' belong to the class of true score.

Dr. Levy begins the formulation of his model with certain definitions and assumptions restated below. I shall refer to Dr. Levy's equations by the arabic numerals of his article and my own by roman. He writes: "Two scores x and y defined in the following way:

$$\left.\begin{aligned} x &= x'+e_x, \\ y &= y'+e_y-f', \end{aligned}\right\} \qquad (1)$$

x is an ability test score and y is an attainment test score; x' and y' are 'true score' components; e_x and e_y are random error components, and f' is "a true retardation component". He further assumes "that all terms on the right-hand side of the equations in (1) are uncorrelated with the exception that

$$x'=y'. \qquad (3)$$

This implies that the true scores of x and y are perfectly correlated and have equal variances". We should explicitly note the inference from the foregoing, that

$$r_{y'f'}=0. \qquad (\mathrm{i})$$

In his introduction Dr. Levy tells us that "the development in this section follows the method used by Gulliksen in his theory of mental tests". Now, although he does not categorically state that he accepts Gulliksen's definition of true score, it is a reasonable presumption that this is his intention, since he nowhere gives a definition of his own. Moreover, his acceptance of the 'usual way' of defining error variance in equations (6) is consistent with this presumption. His equations in (1) do not define the general property of true scores, but are used, as he says, to define x and y. Thus it is on the presumption

---

might then appear. Dr. McLeish apparently supposes that this was in fact how I tested the dull and the backward (*op. cit.*, p. 59); but he has there confused the test used for assessing the intelligence of adults (Group Test 33) with those used for backward children.

that he accepts the following general definition of true score, taken from Gulliksen (3, p. 8), that my subsequent argument is based:

$$T = X - E, \tag{ii}$$

where T is " the gross true score ", X is " the gross raw score ", and E is " the gross error score (random error) ", for an individual, i, whose subscript has here been omitted. Adopting this definition for his test scores x and y, the ' true scores ' will be defined as

$$\left.\begin{aligned} x' &= x - e_x, \\ y' &= y - e_y, \end{aligned}\right\} \tag{iii}$$

And applying these two equations to the re-arranged equations in (1), it is apparent that, while the first equations in each agree, the second equations in each must contradict each other, unless f' is zero. It would be pointless to discuss f' if it were zero. Hence it must be presumed that Dr. Levy's y' does not conform with the true score for y as defined in (iii), although there is no such contradiction for x'. He would thus appear to be using ' true score ' in a different sense when he refers to y' and when he refers to x'. The ambiguity is further increased by his definition (3) which makes possible the substitution of x' (and its derivatives) for y' (and its derivatives). Further exploration of his model would seem to require a meaning to be determined for his y' in terms of true score as defined in (iii). However, to preserve the explicit distinction between his y' and the true score for y, *the symbol* y_a' *will be substituted for Dr. Levy's* y', *and the symbol* y' *will be used with the meaning given in* (iii) *as the true score for* y. The proposed change in the meaning does not create a new model; it is an adjustment of the model to meet the requirement of consistent meaning.

Dr. Levy's second equation in (1) may now be restated as

$$y = y_a' + e_y - f'. \tag{iv}$$

Applying the re-arranged equation in (iii) to eqn. (iv) and re-arranging terms, we find

$$y_a' = y' + f'. \tag{v}$$

It follows, since the correlation of a sum with its constituent part cannot be zero (except in certain non-relevant cases), that

$$r'_{y_a'f'} \neq 0. \tag{vi}$$

Under the condition of consistent meaning for true score, (vi) contradicts (i), since y_a' is Dr. Levy's y' by defined substitution. It is apparent from (v) that f' cannot be expressed in terms of components of y alone, when these are defined as in the second equation in (iii). It is necessary to invoke definition (3) to find an expression for f' in terms of components of x and y. Definition (3) restated becomes

$$y_a' = x'. \tag{vii}$$

Substituting for y_a' in (v) and re-arranging, we find

$$x' - y' = f'. \tag{viii}$$

It has so far been demonstrated that in their original form Dr. Levy's equations did not maintain a consistent meaning for true score, and that when a consistent meaning was fitted to them his other premisses were contradictory. In so far as his subsequent equations are dependent upon the inconsistencies demonstrated in his premisses they are false. However, for the purposes of this argument it will suffice to demonstrate that his eqn. (7), which represents a crucial step in his formulation, is incorrect. Equation (7) states that

$$\sigma_{f'}^2 = r_{yy}\sigma_y^2 - r_{xx}\sigma_x^2. \tag{7}$$

Even on first inspection eqn. (7) looks odd, since $r_{yy}\sigma_y^2$ and $r_{xx}\sigma_x^2$ are the usual expressions of true score variance for x and y in terms of their respective reliabilities and raw score variances, and since earlier Dr. Levy states that the true score variances of x and y are equal, eqn. (7) appears to state that $\sigma_{f'}^2$ is zero. However, a detailed demonstration that (7) is incorrect is necessary.

It can be shown that

$$\sigma_{y'}^2 = \sigma_y^2 - \sigma_{e_y}^2, \tag{ix}$$

and that it derives its statement of the relation between variances from the second equation in (iii). Equations (6) state that

$$\left.\begin{aligned} \sigma_{e_x}^2 &= \sigma_x^2(1 - r_{xx}). \\ \sigma_{e_y}^2 &= \sigma_y^2(1 - r_{yy}). \end{aligned}\right\}. \tag{6}$$

The equations in (6) are not derived from Dr. Levy's premisses, but are, as he says, the usual definitions of error variance in terms of raw score variance and reliability of the respective tests. Applying the second equation in (6) to (ix), we find

$$\sigma_{y'}^2 = r_{yy}\sigma_y^2. \tag{x}$$

Similarly, applying (6) to the re-constituted (5), we find

$$\left.\begin{aligned} \sigma_{x'}^2 &= r_{xx}\sigma_x^2. \\ \sigma_{y'_a}^2 &= r_{yy}\sigma_y^2 - \sigma_{f'}^2. \end{aligned}\right\} \tag{xi}$$

Deriving an expression for the variance of f' from (viii) in the usual manner, we obtain

$$\sigma_{f'}^2 = \sigma_{x'}^2 + \sigma_{y'}^2 - 2r_{xy}\sigma_x\sigma_y, \tag{xii}$$

which after substitution for $\sigma_y'^2$ and $\sigma_y'^2$, in accordance with (x) and the second equation in (xi), becomes

$$\sigma_{f'}^2 = r_{xx}\sigma_x^2 + r_{yy}\sigma_y^2 - 2r_{xy}\sigma_x\sigma_y. \tag{xiii}$$

It is clear that (xiii) differs from (7) and that in particular (7) does not include the necessary component attributable to the correlation between x and y. By this omission alone eqn. (7) is demonstrated to be incorrect. Dr. Levy has been able to derive his eqn. (7) by having two contradictory estimates of the true score variance for y in the same equation (the second equation in (5)). Thus he asserts that $\sigma_{ya}'^2$ (his $\sigma_{y'}^2$) is the true score variance for y, but by deriving $r_{yy}\sigma_y^2$ from the application of (6) to (5), he obtains a second contradictory estimate of the true score variance for y. A comparison of eqn. (x) and the second equation in (xi) reveals the contradiction clearly.

The conditions requisite for expressing 'true retardation' in psychometric terms are contained in eqns. (iii) and (viii), and the variance for true retardation is given by eqn. (xiii). Thus the psychometric formulation of retardation is in no way dissimilar from that for differences between scores on any other pair of tests.

In view of the controversy that has arisen over the relation between ability and attainment it is interesting to speculate how Dr. Levy was led to state his premisses in the inconsistent form I have indicated. His definition (3) provides a clue, since it allows us to substitute the 'true score' underlying the ability test for the 'true score' underlying the attainment test. Now one of the central propositions of current retardation theory which my colleagues and I sought to challenge was the statement which prescribes the use of a score obtained with a test of ability as if it were an estimate of what the score for attainment *might* have been had attainment not been retarded. This seems to be the principle that Dr. Levy had in mind when constructing his model, since it is just such a model to which that principle would have led.

Those of us at the National Foundation for Educational Research who put forward our views in 1956 and 1957 were by no means guilty (as Levy supposes) of confusing 'retardation theory' with 'the numerical expression of it'. Indeed, we recognized that some kind of theory was needed in order to fit the data derived from tests of ability and attainment, namely, a theory which would no longer condone the practice of combining ability and attainment scores in a way predetermined by the assumption that the ability score was an estimate of 'ideal' attainment. We argued (*a*) that the problem of using ability and attainment scores to predict improvement in attainment after remedial treatment should be dealt with by operational approach; (*b*) that backward children should not be deprived of remedial treatment solely because the difference between ability and attainment scores was not on the side of ability; and (*c*) that a revised concept of 'readiness' should replace for teachers the concept of retardation as they understood and applied it.[1]

Dr. Levy suggests that we can consider the score for an intelligence test as "a short term estimate of how well a child is likely to perform new tasks". This is an impeccable statement so long as the idea of prediction is retained; but the use of the intelligence test, where retardation is expressed as a difference between scores with intelligence and attainment tests, implies a good deal more than prediction. We do not wish to predict present where retardation is expressed as a difference between scores on intelligence and attainment, we should use the best predictors, empirically determined. The further implication of Levy's theory is that the intelligence test score measures what the attainment score *might have been* or *ought to be*. The authority for this cannot come from predictive studies, since no criterion exists on which they can be based. The notion is apparently derived from the developmental theory of retardation. It certainly does not arise solely, as Dr. Levy appears to suggest, from the relative 'stability' of the intelligence test score, even if this were self-evident, since the relevance of intelligence to attainment—and (what is probably more important) the relevance of the *particular kind* of 'intelligence' measured by the test employed—must surely affect the issue.

Finally, may I briefly re-state my attitude towards retardation theory and practice?

(1) I agree that all children may be thought of as retarded to an indeterminate degree, though I find little practical use in such a notion, which plainly leaves out of account the value systems by which retardation is judged.

(2) I reject the use of a difference between scores for ability and attainment tests as a measure of the degree to which a child's attainment falls below what it *might be* or *ought to be* or *could be*.

(3) I agree that combinations of ability and attainment test scores may yield prediction of improvement which are superior to predictions furnished by either taken separately; and I welcome systematic studies of the problem, particularly along the lines adopted by Phillips [5].

M. A. BRIMER.

REFERENCES

[1] BRIMER, M. A. (1956). The relationship between ability and attainment. *N.F.E.R. Bulletin*, VIII, 29–33.

[2] BRIMER, M. A. (1957). Ability and attainment—a discussion of some controversial issues. *N.F.E.R. Bulletin*, X, 25–28.

[3] GULLIKSEN, H. (1950). *Theory of Mental Tests.* New York: Wiley.

[4] MORRIS, BEN S. (1956). The relationship between ability and attainment: some reflections. *N.F.E.R. Bulletin*, VIII, 34–36.

[5] PHILLIPS, C. J. (1961). On comparing scores from tests of attainment with scores from tests of ability to obtain indices of retardation by different ratios. *Unpubilshed Thesis*, University of Birmingham, Institute of Education.

[6] PIDGEON, D. A. and YATES, A. (1956). The use of tests in the classroom: an examination of current test theory. *N.F.E.R. Bulletin*, VIII, 24–28.

[7] PIDGEON, D. A. and YATES, A. (1957). Ability and attainment: a further note. *N.F.E.R. Bulletin*, X, 22–24.

---

[1] Our articles were intended primarily for teachers. Whether or not reputable psychologists actually recommend or practice the determination of specific retardation on the basis of a difference between the score for a single attainment test and the score for a single intelligence test, there can be no doubt that teachers did indeed adopt this procedure and are still being trained to do so.

ABILITY AND ATTAINMENT. IV

Mr. Brimer claims that my algebraic derivations are inconsistent. This claim springs from one verbal ambiguity which is reinforced by three erroneous assertions. His first point of attack is that I have not used the expression 'true score' consistently. Since (as Gulliksen defines the phrase algebraically) a 'true score' is simply a score free from error, the two 'true score' components in y are properly described as 'true'. Any ambiguity I may be guilty of is in calling one component 'true *attainment*', when in fact it is only one of the 'true' *components* of attainment. What this true component is should be clear from eqn. (3) in my original article (p. 138). At the worst then this is a temporary verbal ambiguity, and does not lead to algebraic inconsistency.

Mr. Brimer's principal assertion is that his eqn. (xiii) differs from my eqn. (7). Now, I accept all but one of his equations; and this does not affect the present argument. It will be seen that his eqns. (xi) prove my (7) when the right-hand sides are equated according to (vii), which is a restatement of my (3). A little care is needed here to avoid confusing his y' with mine. Since his (xiii) is identical with my (9), he thus asserts in effect that (7) 'differs from' (9). It is not clear to me what 'differs from' means in this context, because eqns. (7) and (9) are uniquely solved by the result in (10), and in this sense are consistent. That of course assumes that by 'differ' Mr. Brimer means 'are inconsistent'; but perhaps he merely means that they 'look different'.

A second erroneous assertion is that my (7) should 'include the necessary component attributable to the correlation' between x and y. If the equations from which (7) is derived are arranged so that x and y are not multiplied, there will obviously be no component directly attributable to the correlation between x and y. Gulliksen makes extensive use of this elementary point in his early chapters.

A third erroneous assertion is shown by the one equation I do not accept. Brimer's eqn. (vi) is a denial of one of my definitions shown in (v) which he asserts is a 'non-relevant' case. Why this is so is by no means clear. If he wishes to propose a model with different assumptions, he is of course free to do so. But he obviously cannot show that my model is algebraically inconsistent merely by asserting that an assumption or definition is 'a non-relevant case'. I feel that 'Non-relevant to what?' is a fair question.

Finally, Mr. Brimer's remarks about the equality of the true score variances—remarks which are to be found above his eqn. (ix)—are the consequence of the original verbal ambiguity. Since the reliability coefficients enter my equations through substitution for the error variances, components which are not in equation here, there is no algebraic proof that I equated y_a' with Brimer's y'.

His concluding qualitative comments deserve a reply; but I will make it as brief as possible. First, I maintain that the quotations (given on p. 141 of my article) from the writings of Mr. Brimer and his colleagues give sufficient indication that they confuse 'retardation theory' with 'the numerical expression of it'. Secondly, I should not wish to dispute who is, and who is not 'reputable'; the remark on the same page arose out of a misunderstanding in the preliminary correspondence between the editors and myself. Finally, I would emphatically suggest that the problems of measurement, which must of necessity be associated with the 'prediction of improvement' mentioned by Mr. Brimer, will involve precisely the same problems as the estimation of retardation.

Mr. Peaker appears to have raised many important points in his discussion with Pidgeon and Yates which are unfortunately not evident in their articles and notes on this subject. Their original (1956) article, in which they acknowledge Mr. Peaker as their stimulus, is almost exclusively concerned with the shape of a distribution and the 'discovery' that the number of children whose superior attainments rise above the mean is fairly large. A subsequent anonymous note (1958) refers to 'instances' in which two normally distributed variables do not have a joint bivariate normal distribution, but says nothing about Mr. Peaker's 'self evident geometric fact' which he encouraged them to demonstrate. Without doubt, had his views been expressed in print before now, much of the controversy would have been avoided.

There is only one small point upon which I wish to contest. He points out that by modifications of the scoring system all children could be made to appear 'under-achievers' or 'over-achievers', and that the usual convention should produce approximately equal numbers above and below the mean. He must surely realize that this was understood. My purpose in writing my paper was to see how well one could develop a more rational numerical expression for the idea that all children are, to a greater or lesser degree, retarded. This surely is the idea that we imply when we talk of 'capacity' limiting 'content', as in the passage quoted by Mr. Peaker and in that quoted by me at the bottom of page 141. Thus, while he and I agree on the really important issues, I do not believe that the choice of convention is anywhere near as arbitrary as he would make out.

PHILIP LEVY.

Vol. XVI
Part I
The British Journal of Statistical Psychology
May
1963

THE USE OF ELECTRONIC COMPUTERS IN PSYCHOLOGICAL RESEARCH

Critical Notice

Computer Applications in the Behavioral Sciences. Edited by HAROLD BORKO. New Jersey : Englewood Cliffs, 1962. Pp. xvi + 633.

The idea of a mechanical aid to computation must be at least as old as the Egyptian abacus, described so vividly by Herodotus. But the first effective digital computer—a desk adding-machine—was constructed in 1642 by a lad of nineteen to help his father in his work as an officer of the French customs: Blaise Pascal's adding-machine can still be seen in the *Conservatoire des Arts et Métiers*. Thirty years later Leibniz exhibited before the Royal Society in London an invention of his own, working by 'stepped wheels', which would not only add but multiply. At the Grand Academy of Lagado Gulliver watched a mechanical device consisting of rotating cubes in a rectangular frame, operated by forty iron handles, and occupying 'the length and breadth of a large room'; it could be used, we are told, to generate new and original statements, and was apparently a composite caricature of Napier's 'bones', Raymond Lully's 'reasoning machine', and Leibniz's proposal to reduce logic to algebra.

And yet even as late as the middle of the nineteenth century the tables in the *Nautical Almanac* were still being calculated with pen and paper by the officials of Greenwich Observatory; as a result they were so out of date, and contained so many inaccuracies, that no experienced navigator would use them. Early in that century Charles Babbage (the wealthy and brilliant son of a banker, elected to the Royal Society at the age of 22 and at 36 to Newton's Chair at Cambridge) had realized that, if only such computations could be mechanized, they could be carried out with far greater exactitude and speed. He began by constructing several 'difference engines' (portions of one are on view at the Science Museum, South Kensington); and then in 1833 hit on the idea of an 'analytical machine', as he called it, which was to form a general purposes computer. He spent over £10,000, partly from his own fortune and partly from funds supplied by the British Government. But, owing to the difficulty of cutting the mechanical parts with sufficient precision, none of his models was ever completed. The fragmentary accounts that Babbage himself has left, couched in a symbolism all his own, are almost unintelligible. The best description both of his methods and his machines we owe to another remarkable personality.

In 1815 Lord Byron married a lady whom he christened his 'Princess of Parallelograms'. In *Don Juan* she figures as Donna Inez; and the first thirty cantos describe the folly of a poet who was rash enough to "wed a walking multiplication table". Just before "they were for ever parted" a girl was born—

Ada, sole daughter of my house and heart,
A child of bitterness, though born in love.

When eighteen Ada began studying under Augustus de Morgan, Professor of Mathematics at University College, London, and was taken to see "Mr. Babbage's engine". "Young as she was," says Mrs. De Morgan, "Miss Byron at once saw the beauty of the invention and understood its working". Shortly afterwards she married the first Earl of Lovelace,[1] but continued working with Professor Babbage. From her account, with its useful mathematical notes (illustrated by charts and diagrams, and incidentally correcting a mathematical slip of Babbage himself), it is clear that the principles of the 'analytical engine' were essentially those of the twentieth century digital computer. To feed in the figures Babbage planned to use the punched cards that had already been devised by a French weaver, Joseph Jacquard, for controlling his looms. A few years later another ex-pupil of De Morgan's, Professor Stanley Jevons, a great admirer of Babbage, constructed a 'logical piano' which would solve quite elaborate problems in deductive logic: (see his *Principles of Science*, 1874, and refs.).

---

[1] When in a broadcast talk on computers, I mentioned the Countess of Lovelace and her mother and daughter as examples of a unique family in whom mathematical ability seemed to be sex-linked to the female members (a pedigree overlooked by Galton in his studies of hereditary genius) I received a letter from one of her descendants saying that several living members were similarly gifted and that one was now working for a well-known firm manufacturing modern computers.

The use of punched cards for recording, storing, sorting, and analysing data was adopted by Dr. Herman Hollerith, of the United States Census Bureau, to deal with the vast amount of information collected during the U.S. census of 1890: with the methods used for the 1880 census the job would have lasted until after the next decennial census. Similar methods were adopted for the British census of 1911. Hollerith's ideas were taken up and developed by the International Business Machine corporation (IBM). However, a machine which fully exploited Babbage's principles was not constructed until more than a century after he and Lady Lovelace had formulated them. This was the 'automatic sequence controlled calculator', commonly known as 'Mark I', designed by Professor Aiken of Harvard in conjunction with IBM, and in operation since 1944. It was essentially an electro-mechanical computer.

The first all-electronic computer was the Electronic Numerical Integration and Calculator (ENIAC), built at the School of Engineering, University of Pennsylvania, in 1946. It was intended primarily to solve differential equations of the type used in calculating the trajectories of bombs and shells, but was capable of most of the elementary arithmetical operations. It had no moving parts; and consequently worked about a thousand times as fast as Mark I. Still more ingenious models were planned by Professor von Neumann and his co-workers at Philadelphia, and affectionately christened EDVAC, UNIVAC, and MANIAC. Their distinctive features were, first the substitution of the binary number system (the conversion from the decimal system being carried out by the machine) and secondly the internal storage of the programme of instructions. A similar machine, with an exceptionally large memory, was the Electronic Delay Storage Automatic Calculator (EDSAC) built under the direction of Dr. M. V. Wilkes at the Mathematical Laboratory, University of Cambridge. This was in fact the first of the binary machines to come into actual operation (1949).[1] In the same year F. C. Williams and T. Kilburn, of Manchester University, published a remarkable paper showing how an ordinary cheap cathode ray tube could be used to store information by means of a charge put on to the phosphor screen by an electron beam and to read it by turning the beam once again to the same place. About 2,000 binary digits could thus be stored on a single 5-in. cathode ray tube with an access time of only 10 to 20 microseconds. Whirlwind I, designed and built at the Massachusetts Institute of Technology, reduced the access time to 1 microsecond. This was the forerunner to the computers used in the Semi-Automatic Ground Environment system (SAGE), sponsored by the U.S. Air Force, and designed to process at top speed vast quantities of radar and other data. SAGE is said to be the largest computer, or rather network of computers, in the world today.

The modern computer seems destined to play the same part in the development of the newer and more complex sciences that the telescope and the microscope played in the development of the older sciences. The capabilities of existing machines may be roughly indicated as follows. Since multiplication is the dominant operation, speed is commonly expressed in terms of the number of multiplications (with about 10 decimal digits) carried in 1 second. The present day speed of the best University machines is about 4,000 per sec. The NORC built for the U.S. Navy runs to 33,000 per sec. (With pencil and paper *one* such multiplication takes about 4 minutes; with a desk machine, 15 secs. or thereabouts).

---

[1] Cf. M. V. Wilkes, *Automatic Digital Computers* (1956).

The word memory is usually about 4,000 words, but the newest models run to 8,000 or more. They have an input rate of about 30,000 per minute and an output of about 18,000 per minute. The Mercury can handle 125 variables: the Atlas more than 200.

At the moment of writing there are twelve British Universities which have installed an electronic computer—a Mercury, Deuce, or Pegasus—for purposes of general research. At the University of London Computer Unit a Ferranti Mercury has now been in operation for over four years. During that period, I understand, nearly 500 hours of machine time have been devoted to statistical calculations carried out for social scientists (economists, sociologists, psychologists, and educationists). Up to the present such work has been limited mainly to the calculation of correlations and multiple regressions and analysis into principal components. But plans are already under consideration for extending the types of programme available; and in the next few months an Atlas machine is to be installed, and should be fully operative early next year. The Atlas (developed by Ferranti Ltd. in conjunction with the Electrical Engineering Department of the University of Manchester) is one of the largest automatic data-processing machines in the world. The Mercury Autocode can still be used (as well as programmes based on the Atlas's own system); but the number of possible variables is increased from 125 to over 200.[1]

Considering the active part played by British mathematicians in the design of calculating machines it is surprising to find how little use has been made of such equipment by British psychologists. They write eagerly about cybernetics, machines and minds, and electronic models of the brain: but hitherto they have made little practical use of such automatic aids. As early as 1949 one of my own research students, Dr. Max Hamilton, analysed data of his own with the help of the Mallock machine at Cambridge; and a contemporary student, Charles Wrigley, was one of the first to use an electronic computer for purposes of factor analysis: (see this *Journal*, V, 1952, pp. 105f. and refs.). However, nearly all Professor Wrigley's work has been carried out in the United States. A review of the possibilities of digital computers was published a little later in this *Journal* (VIII, 1955, pp. 59f.); and quite recently one or two notes on the subject have appeared in the Society's *Bulletin*. At University College in addition to several large scale investigations already carried out, Dr. Charlotte Banks and her co-workers are planning to make extensive use of the University's computer to process an immense amount of data collected during a prolonged investigation into juvenile delinquency, as well as for factorial work. It is, however, only within the past twelve months or so that research students in psychology and education have started to publish statistical studies, mainly factorial, based on calculations carried out by an electronic computer. Unfortunately the commonest practice consists in collecting large arrays of test-results, often without any prearranged plan or problem clearly in mind, and then handing over the data to a computing unit to calculate the correlations and perform the factorization and the rotations; the student then contents himself with interpreting the factors thus extracted, usually in terms of his particular preconceptions. The consequence is that, more often than not, he knows less about the rationale of the statistical techniques involved than his predecessors who carried out their own computations with the aid of a slide-rule or a desk machine. This, I fancy, is very largely because there has been no elementary textbook, written by psychologists for psychologists, explaining in comparatively elementary terms the nature, possibilities, and methods of programming and utilizing the various machines now available.[2]

---

[1] I am indebted to Dr. Benedict Nixon, of the University of London Computer Unit, for the above information as well as for many of the comments quoted below.

[2] Both Dr. Nixon and Dr. Ehrenberg (*The Statistician*, III, 1962, 3) have commented on the fact that, as compared with research students in physics or engineering, "psychological investigators appear to have little insight into the underlying mathematical procedures", and ten to regard the computer as " a heaven-sent device for analysing complexes of 100 variables instead of the 10 of hand-computation days ".

Such a textbook is now provided by Dr. Harold Borko, who has edited a volume surveying the whole subject. Dr. Borko is an instructor in psychology in the University of California and a member of the research staff of the System Development Corporation dealing more particularly with what he calls " human factors science ". His twenty-two collaborators consist of specialists actively engaged in behavioural researches which involve the use of an electronic computer.

The book is divided into three main parts. The first two are written entirely by the editor. Part I explains what computers are, and how and why they have developed. " A computer ", we are told, " is a machine that is able to calculate and perform sequences of arithmetical and logical operations in accordance with the prearranged programme of instructions." However, as we shall see, it has much wider uses than this. Its special advantages are its amazing speed, its increased precision, and its economy of mental and manual labour: since it can be programmed, it eliminates the need for human intervention at each successive step.

After a few introductory pages Dr. Borko begins by disposing of what is commonly the first question asked by the naïve student—" does an electronic brain really think ? " As C. E. Shannon has observed, " The answer depends on how we define thinking: from a behaviorist standpoint it certainly *acts as though* it were thinking ". In point of fact, says Dr. Borko, " the computer only performs prescribed physical operations which the human operator *correlates* with thinking. Machines manipulate symbols; it is the human interpreter that thinks ". Most of Part I, however, is concerned with the detailed history of the main stages by which the modern electronic computer has developed. Writing primarily for American readers, Dr. Borko confines himself mainly to American work and American machines; and it is for this reason that I have ventured to describe a little more fully the contributions made by British mathematicians and scientists. He ends the section with a brief account of the characteristics of present-day computers and their main uses.

Their uses are in fact far more varied than the layman commonly realizes. Slightly modifying Dr. Borko's scheme, one may say that (1) their widest use is for the processing of business data for large corporations and governmental offices (e.g. to deal with production, sales, accounts, etc.); (2) their more recent applications have been in the actual control of industrial processes (e.g. spinning and weaving, regulating machinery by feed-back or forward-feed, sorting, inspecting, and even assembling): these two broad types of application roughly correspond to what is known as ' automation ';[1] (3) their most impressive use at the moment is for the U.S. military establishments; and much of the money spent on their development has come from the U.S. Army, Navy, and Air Defence Command; (4) their most obvious and immediate application is to undertake highly complex mathematical computations, particularly those of a statistical type: and since, as Russell long ago maintained, " formal logic is the same thing as mathematics ", the phrase ' mathematical computation ' must here be taken to include what would formerly have been called ' logic inference ', whether deductive or inductive, demonstrative or probabilistic; (5) a group of applications of special interest to pure and applied psychologists, particularly those working in the field of individual and social psychology, consists in the employment of computers as working models which will simulate complex processes or systems. Although included in his book, this last group is not separately classified in Dr. Borko's preliminary list. The examples most familiar to the British psychologists are Grey-Walters' ' electro-mechanical tortoises ', and the more recent ' working models of the brain '; in engineering perhaps the most striking applications are the ingenious models simulating the probable behaviour of new aircraft before they are actually constructed by calculating various flight parameters—speed, gravitational forces, strength of materials, manoeuvrability, etc.

---

[1] Apparently the word was first coined by D. S. Harder of the Ford Motor Company, in 1947. But it seems also to have been used by J. Diebold of the Harvard Business School about the same time: (see his book on *Automation: The Advent of the Automatic Factory*, 1957). For the more general techniques—making a process or system automatic—the older and longer term, ' automization ' should, I would suggest, still be retained.

Part II, though comparatively short, forms one of the most useful sections in the whole volume. It begins with detailed descriptions of the 'functional components of a data-processing system'—the input and output devices, and the units for storage, for control, and for arithmetical computation. Each is clearly explained, with the aid of diagrams and illustrations. The next chapter deals with the representation of data by the binary system, by Boolean notation, by 'flow diagrams', and by punched cards or tapes. This is the stage at which British students of psychology seem most often to get bewildered. Yet for those who have picked up the simple logic of dichotomous classification as it appears in elementary factor analysis, the underlying principles should already be familiar.[1] To take an elementary example, suppose we want to select the brightest candidate out of a group of three—X, Y, and Z say. There are six possible orders; but the solution can be obtained by asking only two or at most three questions. Thus the flow diagram might read: "(i) Is $X > Y$? (A) If 'Yes, $X > Y$' then (ii) Is $X > Z$? (a) If 'Yes', print X; (b) If 'No', print Z. (B) If to question (i) the answer is 'No, $Y > X$', then (iii) Is $Y > Z$? (a) If 'Yes', print Y, (b) If 'No', print Z" (where $>$ means X's score is higher than Y's, and so on). Every research student should, in my view, be able to analyse his main problem, and the various hypotheses he wishes to confirm or refute, into a logical scheme of this kind. To convert the scheme into a conventional 'flow diagram' a specialist—the programmer for the particular machine which is to be used—will doubtless be required, and he may need to consult the student. The programmer will probably begin by translating the task which the student wants the machine to carry out into what is known as a 'problem-oriented language'. This should be intelligible to the human analyst, once he has learnt the vocabulary. Perhaps the most effective and widely used language is that known as FORTRAN (a portmanteau term for 'formula translation'). This was intended primarily for the more recent models built by IBM, but can also be used for older models and for a good many machines constructed by other firms.[2] The 'language' used by the University of London Computer Unit—the 'Mercury autocode'—is very similar in principle, but has a slightly different vocabulary.[3] 'Fortran' will then have to be translated or coded into a language that can be understood by the machine. Here again there may be two sub-stages. 'Fortran' may first have to be translated into a 'universal computer-oriented language'; that in turn will need to be reformulated in terms of the sequence of the computer operations to be performed, and that may itself require a code peculiar to the type of machine to be employed. Programmes for the commoner computations required by psychologists are now available both in fortran and in the 'universal computer-oriented language'[4]; and with a little helpful instruction most research-students can learn to punch cards or tapes for their own data. A satisfactory planning of the whole research at the outset, however, becomes far more complicated. Hence there should be far more consultation between the psychological research-worker and the programmer at successive stages of the investigation; and to this end it is essential that the research-worker should have some elementary acquaintance with the techniques and the terminology used in programming computers.

---

[1] For a further explanation see my previous discussion of the problem in this *Journal*, *VIII*, pp. 62f.

[2] See International Business Corporation, *General Information Manual: Fortran*, (1961). New York: IBM Processing Division, and D. D. McCracken, *A Guide to Fortran Programming* (1961), New York, Wiley.

[3] For a fuller account see R. A. Brooker, *Mercury Autocode Manual*, (1961), London: Ferranti Ltd.

[4] Dr. Borko and his collaborators include a few typical programmes of both types. But for practical purposes perhaps the best collection is to be found in *Multivariate Procedures for the Behavioral Sciences*, by W. W. Cooley and P. R. Lohnes (John Wiley, 1962). This also uses 'fortran' for the flow diagrams.

In addition every student should try out the methods to be used for himself on a desk machine. The non-mathematical student is all too ready to hand over all calculations to the computer. He is moreover a little apt to suppose that computers, unlike human beings, never make mistakes. More than once I have seen results passed back to, and accepted by, research-workers, which are manifestly absurd even on a first inspection.

Part III, which occupies more than half the volume, deals with typical applications of computer techniques that are likely to be of interest to psychologists and social scientists. Each chapter is written by an expert who is himself actively engaged in original research; and every contributor has evidently been asked to illustrate his special theme with concrete instances, so that the exposition proceeds throughout in accordance with the admirable principle of teaching by example.

For convenience, the whole part may, I fancy, be subdivided into four distinct sections. The opening chapters by Professor Wrigley, E. L. Kelly, and J. C. Lingoes (all now of the University of Michigan) deal with ‘the university computing centre’ and ‘data processing in psychological research’. Dr. Wrigley incidentally contends that every large department of psychology should include a computer specialist. The specialist should at the same time be a member of the computing centre, and might, in the early stages, also serve other departments—sociology, political science, communications, and the like. It is to be hoped that this informative chapter will be studied by the head of every psychological department in the country. Dr. Kelly and his colleagues point out that in the United States, next to the mathematicians and physicists, the university psychologists were among the earliest and most frequent users of computers. Psychologists have to deal with fallible measurements: they cannot, as a rule, plan their experiments so that all variables except one are strictly constant; and the relations that thus arise are highly complex and consequently require one or other of the available forms of multivariate analysis. Indeed, as the authors rightly remark, “many widely used statistical techniques were developed by psychologists themselves”—notably by Galton and his followers. “The potentialities of the computer seem first to have been realized by psychologists at the University of Illinois”; and it is gratifying to find that two former research-students from University College, London, are named as leaders in this movement—Professor Raymond Cattell and Professor Charles Wrigley. Readers of this *Journal* will already be familiar with their work from papers in earlier issues.[1]

The second section deals with specific statistical types of calculation—multiple linear regression, factor analysis, and canonical analysis. The last may strike some readers as a novelty. As the author of the chapter, Dr. P. B. Koons, observes, “not until 1947, when the *British Journal of Psychology* added a Statistical Section,[2] was canonical analysis brought to the attention of the psychological public”. Hitherto the complexity of the calculation has prevented its practical use; but today these can readily be overcome by the use of computers. Multiple discriminant analysis, which has suffered neglect for much the same reason, might, I venture to think, well have been added to the series, together with a few more general techniques such as the analysis of variance and covariance and some of the various classification procedures. The chapter on the all-important topic of multiple regression is by no means as clear as it might be; but the following chapter on factor analysis by Fruchter and Earl Jennings is admirable. Since Pearson's original procedure, from which all the rest have developed, is equivalent to extracting Eigenvalues and Eigenvectors, almost any machine can at once undertake it.[1] Programmes are now also available for rotation by the quartimax and varimax techniques. Factor analysis however, perhaps more than any other method, is a case where frequent consultation between the programmer and the investigator is desirable. In this country most students seem content to allow the programmer to put unity in the leading diagonal—a device which is liable to produce an excess of artificial factors. It is a simple matter to insert and improve communalities, though for this the investigator will usually require to study his results after each successive analysis. But the most urgent need seems to be for a programme to deal with 'group factor analysis' (i.e., the method which allows a 'general factor' jointly with a 'simple structure'), and including the necessary rotations for obtaining 'overlapping group factors'.

---

[1] See the paper published 10 years ago by C. Wrigley and J. O. Neuhaus on ‘The Refactorization of the Burt-Pearson Matrix with the Electronic Computer’ (this *Journal*, V, 1952, pp. 105f.). The work was carried out on the Ordvac (built by the Illinois School of Engineering for the Ordnance Department of the U.S. Army) almost as soon as it was in operation. The comparison with results previously obtained with a desk machine admirably demonstrates the superior range, precision, and speed now obtainable: calculations which formerly took a week or more of full-time labour can be accomplished in 5 or 10 minutes. See also the factorial analysis of the 11 × 11 correlation matrix for emotional traits (*ibid.*, XII, pp. 38f.). For rotations see Wrigley and Neuhaus ‘The Quartimax Method’ (this *Journal*, VII, 1954, pp. 81f. and refs.).

[2] The reference is to ‘Factor Analysis and Canonical Correlation’ (this *Journal*, I, 1948, pp. 96f.). Dr. Koons also cites the papers by Thomson on ‘The Maximum Correlation of Two Weighted Batteries’, Bartlett on ‘Internal and External Factor Analysis’ (both in the same volume), and Wrigley on ‘The Prediction of a Complex Aptitude’ (*ibid.*, V, pp. 93f.)—all of which deal with much the same problem.

The third section may perhaps best be described as dealing with the simulation of processes, psychological and neurological, *within the individual.* Ross Ashby discusses 'the simulation of a brain', J. T. Culbertson the processes in 'nerve nets'; other writers describe 'computer simulation' of perception, learning problem-solving, 'language behaviour', and musical composition. Here may be included the chapters on 'automated teaching' and on the use of computers to generate visual displays. The last section might be described as dealing with the simulation of processes *in social organizations*—e.g., 'productive systems', 'governing systems', and 'business organizations' generally—and the simulation of international relations and interactions.

Dr. Borko rounds off the whole with 'a look into the future'. As regards hardware design he foresees processes that now take milliseconds or microseconds being accomplished in a 'nanosec' (a billionth of a second). Even more important is the need for increasing storage capacity. At the moment, however, by far the slowest and most troublesome part of the procedure is the preparation of programmes. A development that will doubtless appeal most strongly to the psychologist is the attempt to design equipment which will automatically scan a page of print or typescript, and convert the contents into the requisite code. Meanwhile what would seem to be still more urgent is the production of a universal, or at least a more broadly standardized, computer language, applicable to a far wider variety of machines. At present a variety of specialized codes are needed because different computers and different programmers adopt their own peculiar vocabulary.[2] And of course there are numerous experimental and mathematical procedures used, or being tried out, in psychology for which at present no programme whatever exists. Here is a task where the psychologist himself must take the initiative.

---

[1] For the analysis of a correlation matrix into its principal components (Pearson's Method of Principal Axes or 'Weighted Summation') most programmers in this country rely on the Oxford Programme for calculating Eigenvalues and Eigenvectors. For a fuller account of American usage see C. Wrigley and J. O. Neuhaus, 'The Use of an Electronic Computer in Principal Axes Factor Analysis', (*J. Educ. Psychol.*, XLVI, 1955, pp. 31–41, D. R. Saunders, A Computer Program to Find the Best Fitting Orthogonal Axes for a given Hypothesis, *Psychometrika*, XXV, 1960, pp. 207-210, and H. F. Kaiser, 'The Application of Electronic Computers to Factor Analysis', *Educ. Psychol. Meas.*, XX, 1960, pp. 141–151. I may add that the issue of an *Statistician* for 1962 (XII, iii) was also devoted to factorial techniques.

[2] Dr. Nixon is developing " a single, general and powerful programme which will provide for all and any of the commoner forms of multivariate analysis". It will be, " so far as possible machine-independent, i.e. not tied to the peculiarities of any particular machine, and fully automatic in operation "; the programmes now in use, though greatly improved as compared with those available 4 or 5 years ago, still require the services of a specially trained operator. Dr. Nixon calls his method Mulvaran, and hopes that it will be available early next year.

But, on looking back over the whole volume, what will, I suspect, impress the British reader most vividly is the new and expanded view he thus gains of what a computer is and what it can do for psychology. The term ' computer ' is itself misleading. The ' electronic data-processing machine ' is not just an arithmetical drudge, or even a novel logical reasoner. Such appliances are not *mere* calculators; they are manipulators of symbols. The manipulation may solve arithmetic, algebraic, or logical problems: but it can do much more. The symbols may represent many other things besides numbers, algebraic variables, or formal propositions: they can represent almost any fluctuating element in any type of system. Hence, as we have seen, a so-called computer may readily function as a model which reproduces (no doubt with a good deal of simplification) the conduct of a community or the behaviour of a brain. Consequently the psychologist should think of a computer, not merely as a highly elaborate substitute for his desk machine, but as an aid to experimentation. For this reason, I would venture to urge that, not only every statistical psychologist, but every psychological student, should make himself familiar with this volume.[1]

CYRIL BURT

---

[1] Since the above was written, the British Psychological Society has arranged a Symposium, on ' Possible Uses of Electronic Computers in Psychology ' as part of its 1963 Annual Conference. The contributors included Benedict Nixon (University of London Computing Unit), R. A. Brooker (Department of Electrical Engineering, University of Manchester), K. A. Redish (Computing Services, University of Birmingham), F. W. Warburton and G. M. Forrest (School of Education, Manchester), and H. J. Hallworth (Education Department, Birmingham). We hope to publish some of the papers in a later issue. It is instructive to note that, whereas, in the early days of statistical psychology, factorial and other techniques were developed chiefly by students of individual psychology working in the field of education, so today it is British educational psychologists who are most active in adopting or adapting modern computer techniques. In the course of the discussion Dr Max Hamilton, who is probably more experinced in work with computers than any other psychologist in this country drew special attention to the difficulties under which British investigators labour at the present time. If the psychologist has not only to learn how to make his own programmes and to cope with the vagaries of the machine, but also to wait several weeks before it is available, the advantages gained by the speed of the computer may be entirely lost; and if the research itself is exploratory, it is impossible to expect the results desired to be produced at one fell swoop.

May I draw the reader's attention to a volume recently published under the slightly misleading title of *Management and the Computer of the Future* (edited by Martin Greenberger and published by the MIT Press) ? During its centenary celebrations the Massachusetts Institute of Technology organized a series of talks and discussions on the future of the computer—including such topics as the place of the computer in the University, new methods of programming, the simulation of human thinking, decision making, and the like. The book is written by experts for experts; but should nevertheless be of interest to the statistical psychologist. Unfortunately there is little reference to recent developments in this country; and it is to be hoped that a similar volume will be prepared and published over here.

BOOK REVIEWS

Combinatorial Chance. By F. N. David and D. E. Barton. London : Charles Griffin & Co., 1962. Pp. x + 356. 63*s*.

" Bernard Bauhusius, some time before 1617, discussed the number of possible arrangements of the words in the line

Tot tibi sunt dotes, Virgo, quot sidera caelo ".

(' Thy gifts, O Virgin, are as numerous as the stars in the sky '). He concluded that, if we reject those that are derogatory to the Virgin, the number of arrangements would be 1022, " which, as Puteanus points out, is also the number of stars in Ptolemy's ' Catalogue ' ". There are eight words in the line; hence the total number of possible arrangements is 8! = 40 320; " so that, if we put down the words in a random order, the probability that they will convey Glory to the Virgin is 1022/40 320 = 0·02535 ", or 39 to 1 against. This is one of the many entertaining examples with which Professor David and Dr. Barton commence their exposition of what develops into a most complicated study with many unexpected implications, practical as well as theoretical; and it may serve to illustrate the vast amount of research which the authors have undertaken, not only into the mathematical analysis of the problems involved but also into the long and curious history that lies behind them.

They start with a chapter on ' elementary combinatorial method ', and then proceed in turn to ' definitions ', ' probability-generating functions ', ' moment-generating functions ', ' expectation techniques ', and ' the distribution theory of runs '. This brings them to the problems of ' derangements and matching '. The ' matching method ' is a device which has been employed in a wide range of psychological investigations—from experiments on telepathy to the evaluation of complex judgements on individual personality. Yet, oddly enough, it is hardly ever discussed in the stock textbooks on psychological statistics. Mathematicians, from the seventh century onwards, have been interested in calculating the betting odds which arose in the ' game of encounters ' (*recontre*)—a pastime which was highly popular with early French card-players and has survived in the children's game of ' Snap '. Dr. David and Dr. Barton deal systematically with the whole topic, starting almost from first principles, along the lines adopted by Whitworth—an adaptation of De Moivre's method of successive inclusion and exclusion. The expressions obtained from a rigorous treatment unfortunately prove far too complicated for practical use; and accordingly the authors go on to consider the applicability of various simpler approximations.

Considerable space is devoted to another set of problems that constantly crop up in psychological investigations, namely, those arising from the use of ' ranks ' or ' grades ', i.e. what teachers usually call ' order of merit '. They are problems which have been intensively studied only during the last fifty years or so. This, as our authors observe, " is surprising, because much of ranking theory as applied to random sequence depends on elementary properties of numbers, which have held a fascination for the pure mathematician for centuries ". " The impetus to study the theory of ranking ", it is said, " came only with the development of psychological research and the realization that absolute measurement of such things as personal preference, for example, was not feasible. Until the advent of Spearman little attention was paid to ranked observations ". Spearman himself tells us that it was Francis Galton who first called attention to the value of such devices. Galton was certainly the first to introduce the word ' Rank ' (spelt with a capital R) as a technical term. Even today, however, much of the theory is still " in a formative stage ". The chapters on the subject begin with the question of ranking criteria designed to test for randomness in a sequence of two or more alternatives and then consider the more important bivariate statistics, particularly those associated with the names of Spearman and Kendall.

The book concludes with an examination of the various ways in which possible departures from random arrangements may be described, and of the resulting effects on the distribution of the criteria—i.e. what are often termed ' non-central distributions '. Here no completely unified theory has been achieved. The various types of ' non-central alternatives ' are grouped into a few broad classes; and each is separately discussed.

Out of the many topics with which the whole volume is concerned, I have mentioned only those which have the most obvious bearing on the work of the statistical psychologist. But everyone who is specializing on research in this field should make a point of closely studying every chapter. The treatment of the issues, complex as they often are, is unusually lucid, and the felicitous choice of illustrative examples makes the discussion delightful to read.

CYRIL BURT

The Behavioral Basis of Perception. By J. G. TAYLOR. New Haven and London: Yale University Press, 1962. Pp. xvi + 379. £3 8*s*.

Mr. Taylor's general thesis " is that all conscious experience is a function of learned behavior, although " (he adds) his " exposition is confined to visual perception ". His immediate aim is " to construct a theory of perception within the framework of the natural sciences " by demonstrating that perception is " a necessary and predictable consequence of the properties of living organisms ". However, his exposition is by no means easy to follow, since, " to avoid ambiguity and longwindedness ", he has decided to express it in terms of " mathematical set theory". As long as he keeps to his symbols his deductions seem fairly cogent. The difficulties arise over his initial postulates and definitions, and his final attempts to identify complex symbolic expressions with observable events or processes. His main argument begins by defining a set, F, which is essentially " a list of the ordered pairs ", q and m, satisfying the condition that m is a conditioned response to q. " For the set of responses, m ," we are then asked to substitute " the set of engrams, e, networks of neural connections." This yields a new set Φ with elements ϕ. Finally, we obtain a third set Π consisting of those elements of Φ which are simultaneous. Thus " Π is a complete description of a momentary state of the system that has emerged from the adaptation of the multistable system, U ", i.e. " of the organism in its material environment."

The system which thus emerges and which is designated Π constitutes, we are told, " a very remarkable set ", and is discussed in some detail both in the opening and in the concluding chapters of the book. According to the definition with which he starts, Π is merely a specific set of engrams. However, Mr. Taylor now " proposes to identify Π with perception ". " If this is a valid hypothesis, then ", so he contends, " the properties of perception should be directly deducible from the properties of Π." He goes on to cite observations and experiments which, he claims, shows that a number of the properties deducible from Π as so defined do in fact characterize the processes of perception. In particular, he tells us, " the set Π describes a collection of sense-data that are in operation simultaneously." How exactly this last statement follows from his assumptions he does not explain. Finally we are asked to " consider an element of Π to be a piece of knowledge about the environment "; and " then ", he says, " the whole set will describe the simultaneous knowing of a large number of facts about the environment; but this is precisely what is meant by being conscious". The last statement, however, seems a definite non-sequitur. For, if an element is merely a " network of neural connections ", i.e. a structural part or " property " of a " living organism ", it is difficult to see how it can also be " a piece of knowledge". At all events the argument cannot possibly be said to prove that perception, as a mode of knowing, is a " *necessary and predictable* " consequence of the properties initially assumed.

Mr. Taylor, however, contends that his theory as thus set forth enables us to dispense with " the disjunction of subject and object that has been a basic axiom for practically all writers on the theory of knowledge". His critics, he says, may attempt to argue that "the sense-data on which all knowledge is based are immediate and unanalysable experiences", i.e. something given, not something deduced. But any such objection "reasserts the dogma of the ghost in the machine"; and this, he holds, has been shown to be completely untenable by "Ryle's brilliant exposure". "Mind" would be something "distinct from the order of nature"; engrams, being physical, are not. Thus, contrary to the common view, it is after all "not impossible to define consciousness in physical terms". By so doing we get rid of non-natural concepts, such as minds, souls, wills and the like, and at the same time are able to explain "the failure of introspection" as a scientific method. Dualism, in short, is banished. Thus, in the end, "the revolt of the early behaviorists against the 'mentalist' assumptions of classical psychology has led to a new assertion of the scientific respectability of consciousness", for, according to Mr. Taylor's demonstration, consciousness itself turns out to be simply a behavioural, and therefore a physical, phenomenon.

This final conclusion, as Mr. Taylor himself implies, seems decidedly paradoxical. And the unprejudiced reader, whether behaviourist or mentalist, will naturally wonder whether there may not be some hidden snag. Nor is it difficult to see where the fallacy lies. Suppose that as a mere male I am puzzled by what at first sight seems the incomprehensible characteristics of the feminine sex; and, as a thoroughgoing monist, I want to solve the problem by "banishing the dualism" which, according to the popular conception, seems to make femininity "something distinct from the order of" masculinity. I begin by defining a set M consisting of all males. I then, by a stroke of originality, "propose to identify women with part of the set M". "If", I argue, "this is a valid hypothesis, I should be able to deduce the properties that characterize females from those that are admitted to characterize males." What are the most conspicuous properties characterizing males? They all have two legs, two arms, two eyes, an upright posture, and the gift of speech. From my hypothesis I deduce that women will have the same properties. And observation shows that they do in fact have all these properties. It would seem therefore that I have now confirmed my hypothesis, and can therefore claim that femininity is "a necessary and predictable consequence" of the male organism as such.

In like fashion, if we have a passion for eliminating dualism, a demonstration of this kind can nearly always be constructed which will apparently abolish it wherever it occurs. It would be equally easy to "verify the hypothesis" that black is white, and so banish yet another form of dualism. The logical error is surely obvious: it is the familiar fallacy of the consequent: "If A (antecedent), then C (consequent); but C, therefore A"—the commonest of all sophisms. The apparent 'demonstration' proceeds by selecting those particular properties that *are* common to both, and silently ignoring the rest. There are, of course, numerous properties peculiar to the phenomenon of perceptual awareness which are *not* properties of physical structure as such, and so far as an engram is a physical structure it cannot claim the property of consciousness: indeed there are many neural engrams which operate quite unconsciously.

There would be nothing illogical in explicitly *postulating* that certain types of physical structure have the property of mediating, or generating, the unique relation of awareness (the most distinctive characteristic of perception) as an 'epiphenomenal' or 'emergent' result. But the connection between the two—the physical structure and the psychical property—would have to be accepted as an arbitrary and empirical datum. Postulates of this kind have frequently been put forward in the past. But, though they overcome certain minor difficulties, they lead to others which appear still more formidable; and thus on the whole they have tended to hinder rather than to assist research. The objections to such a view are well known; and Mr. Taylor's chapter on 'Other Theories' makes no attempt to rebut them. We must therefore conclude that he has failed to demonstrate what he regards as his main thesis.

On the other hand, the body of his book is undoubtedly of great interest to psychologists, particularly those who have specialized in the experimental study of vision. Ross Ashby's theory of 'multistable systems' provides the basic quasi-mathematical concepts in terms of which the issues are formulated and discussed. Clark Hull's *Principles of Behavior* furnish the chief behaviouristic assumptions. And the experimental work has largely been suggested by the recent researches carried out by Theodor Erismann and Ivo Kohler at Innsbruck.

The topics in which Mr. Taylor is mainly interested arise out of the visual perception of space. He starts with two chapters on the development of space perception, and then proceeds to consider the influence of gravity and of parallax, and the general 'expansion' of the visual field. Finally he examines, rather more briefly, the perception of colour and the development of perception through other modalities. Perhaps the most instructive chapters of all are those which report a series of experiments on the effects of wearing various types of 'distorting spectacles', e.g. those which (like Stratton's lenses), not only invert the visual field, but also reverse the directions of right and left—a device which Mr. Taylor considers to be "the most important experiment in the whole history of psychology". The data thus obtained are analysed much more rigorously than has been done before, and considerable use is again made of set theory and mathematical deductions. Dr. Seymour Papert provides a 'mathematical appendix' which enables the effects of wearing such spectacles to be formally calculated. Although none of the results support Mr. Taylor's 'general thesis', the conclusions reached undoubtedly form a valuable contribution to scientific knowledge of the processes involved.

W. L. GREGORY

The Science of Behaviour. By John McLeish. London: Barrie & Rockliff, 1963. Pp. 199, 18*s*. 6*d*.

Dr. McLeish's book has no preface. From the topics selected for discussion, however, it would appear that it is intended, not so much for students of psychology as for students of education, especially those in training colleges for teachers, and for what the dust-cover calls 'the intelligent layman'. Its aim, we are told, is to describe how "the scientific study of the way people behave has made inroads on a territory that was formerly the exclusive province of religion or morality". "The materials on which the book has been based" (Dr. McLeish adds on the last page of all) "are scattered in out-of-the-way periodicals and books, most of which would be unavailable to the ordinary reader. The author has therefore decided not to annotate the text with references to the literature." This, as I think every reviewer will agree, is a most regrettable policy, particularly since Dr. McLeish describes his approach as essentially "polemical". As we shall see later on, he attributes the most surprising views to various living authors; but readers are given no opportunity for checking the accuracy of his accusations.

The book is divided into five main 'parts', the last consisting of a single chapter only. Part I, entitled 'Presuppositions', contends that, as a natural science, psychology rests on three basic assumptions—"simplicity, order, and causation". "Theological categories", which apparently include such concepts as 'mind' and consciousness, are rejected because they "multiply entities which are unnecessary", and "violate the principle of causality." The concluding chapters explain that Pavlov, who is described as an "atheistic monist and materialist", "formulated a coherent and comprehensive theory of human organism as a self-regulating system", with "the principle of determinism as basic". His theory is held to be "sufficiently sophisticated to encompass the complexities of human activity", and therefore forms the groundwork of the book. Dr. McLeish tells us that "the importance of Pavlov for psychology only became apparent as a result of the discussion" (details not given) "which took place in 1950 in the Soviet Union". It now provides us with an "objective psychology freed from the Procrustean framework of introspection and theology." It is true, he says, that Sherrington believed that consciousness as such was not reducible to terms of electrochemical processes taking place in a material brain; but that is because Sherrington's thinking was "vitiated by an underlying theological dualism". To anyone who is familiar with Sherrington such a criticism will sound almost ludicrous; Sherrington based his statement on what he considered to be the facts; and certainly his knowledge of the nervous system was at least as thorough as Pavlov's. He was in fact the first writer in this country to mention Pavlov's work—as early as 1906. Dr. McLeish seems completely to have forgotten that Watson attached increasing importance to the Pavlovian theories in his expositions of behaviourism from 1914 onwards, i.e., long before 1950, and that Woodworth included them in his *Contemporary Schools* (1931) and devoted a whole chapter to them in his text book of *Experimental Psychology* (1938).

Dr. McLeish, however, has yet another surprise in store for us. He recognizes that a good many physicists are chary of talking in terms of "causes" and prefer to speak of "functional relations between independent and dependent variables". Certain quantum physicists have even put forward a principle of indeterminacy. But that, he tells us, "is not regarded favourably by scientists outside a comparatively narrow circle. The majority of scientists would now agree with Leucippus." Why the scientist of today should trust Leucippus rather than Heisenberg is by no means clear. But in any case Dr. McLeish's main purpose remains clear: it is to make out the strongest possible case for retaining the materialistic and deterministic type of monism that was popular among German writers in the nineteenth century and to show how it can form the basis for modern psychology. His brand of behaviourism is therefore quite as extreme as that of Watson and quite as out of date.

The chapters which will chiefly interest readers of this journal are those which make up Part II and deal with "the theory of mental testing". Here the main target for attack is "the statistical basis of the theory" (or rather what Dr. McLeish supposes to be its statistical basis). He begins with an attack on "the Galtonian and Pearsonian methods of mathematical analysis" as adopted and developed by Spearman, Burt and their various co-workers. In the first chapter which opens this part he declares that "the construction and use of intelligence tests is based on a phenomenalist approach which denies the possibility of studying real causal connections . . . Discussion of the nature of intelligence has moved from a behavioural universe to an artificial mathematical universe of vectors and matrices, oblique and orthogonal axes, which merely clouds understanding, and ensures that the mental tester will fail to say anything significant about human behaviour". Here Dr. McLeish reveals at the very outset that, in common with many other non-statistical writers, he completely misunderstands the function of mathematical techniques in psychology. And to represent Burt, who has always been one of the most vigorous opponents of 'phenomenalism', as adopting it as his basis is, to put it mildly, the acme of paradox.

Galton, Burt, Spearman, Brown, and others who interested themselves in "the construction and use of intelligence tests" never denied "the possibility of studying the real causal connections underlying human behaviour" (provided acceptable definitions are given of 'real' and of 'causal'). Galton, from his investigations of the performances of eminent personages, and Burt, from his observations of the behaviour of school children, concluded that the underlying causes were (i) partly genetic and partly environmental, (ii) partly cognitive and partly motivational, and (iii) partly general and partly specific. This threefold classification implied, among other corollaries, the concept of *innate cognitive abilities*, some of which might be *general*. The notion of innate *specific* abilities corresponded to the traditional notion of faculties. But the concept of "an inborn *general* intellectual ability" was a novelty introduced by Galton.[1] Burt accepted both types of ability; Spearman thought " general ability or intelligence " could be made to account for all the variations observed; Brown, Thomson, Thorndike, and later Thurstone believed " special abilities " alone were sufficient. Here were three distinct hypotheses. The problem was how to decide between them.

---

[1] This definition (and the amplification of it, apparently extracted from the Eugenics Society's 'Occasional Paper, Number Two', p. 5) is quoted and criticized by Dr. McLeish at some length. But he makes no attempt to state or refute the "eight converging lines of evidence" by which the concept was supported (pp. 35f.). Apparently the *ipse dixit* of Dr. McLeish and the well

It was to resolve this triangular controversy that statistical techniques were invoked. The problem was plainly one of classification—classification primarily of modes of behaviour (or ' traits '), and secondarily of individuals (or ' persons ') according to their manifestation of this or that type of behaviour. Karl Pearson's method of ' principal components ' provided the necessary multivariable procedure for classifying both ' traits ' and ' persons '. His early paper (1901) was expressed in terms of a rather forbidding type of determinantal algebra. The word ' matrix ' had already been introduced by J. J. Sylvester (at one time Professor of Natural Philosophy at University College) to denote " a rectangular array of numbers from which a determinant can be derived". A ' vector ' is merely a matrix of one row (or column); and from 1901 onwards ' vector analysis ' became a regular technique in the physical sciences. The application of matrix algebra to problems of quantum mechanics by Born and Heisenberg in the 1920's widely popularized the procedure; and a year or two later, W. F. Sheppard, one of His Majesty's Inspectors of schools, best known to statistical psychologists for his table of the normal frequency distribution, showed how Pearson's proofs could be put into an extremely simple form by using a matrix notation (*From Determinant to Tensor*, 1923). Sheppard suggested to Nunn and Burt (both then professors at the Institute of Education) that matrix algebra could be employed to express more simply the methods of component (or ' factor ' analysis). As the present reviewer remembers from their lectures, both were insistent that these statistical techniques had primarily a *verificatory function*—a caution that Dr. McLeish has overlooked: the purpose of such techniques was to furnish an objective procedure for confirming or refuting alternative hypotheses: that is all. The ' latent roots ' and ' latent vectors ' to which Pearson's procedure leads are merely the factor variances and the factor loadings. By applying the usual tests of statistical significance Burt claimed that this method and its results plainly confirmed Galton's double factor hypothesis, and refuted both the single factor theory of Spearman and the simple group factor theory of Brown and Thomson. *Both* the single general factor *and* the group factors were needed. It is true that from time to time those psychologists who have never troubled to get up the elementary algebra required have complained that such methods " cloud the understanding ". But neither they nor Dr. McLeish have deigned to inform us how the issues which arise can be settled if such statistical methods are to be discarded.

" By following this line ", however, so Dr. McLeish assures his readers, " the intelligence testing movement has deprived itself of the possibility of demonstrating the real connection between the functioning of the central nervous system and behaviour and of establishing a significant genetic psychology."[1] This contention is decidedly odd. In point of fact the earlier arguments in favour of a hierarchical scheme of general and specific factors were originally suggested by the hierarchical structure and functioning

---

known joke about the Irish counsel defending a murderer are held to be a sufficient disproof. On the following page he tells us that " the conception of innate intelligence is clearly traceable to Schopenhauer ". This has caused some surprise among my German colleagues. Can Dr. McLeish supply a reference ?

[1] It would seem that Dr. McLeish got no further than page 5 of the pamphlet he quotes. If he had read as far as the section headed " Physiological Interpretation " he would have found a clear statement of " the connection between the functioning of the central nervous system and behaviour " (as sampled by the tests) and a " significant genetic psychology " invoked to account for the way the underlying tendencies are transmitted. The working out of a hierarchy of ' levels of organization ' within the nervous system (which Dr. McLeish attributes to his hero Pavlov) was really due to the earlier work of Sherrington, though the general concept goes back to Spencer and Hughlings Jackson.

of the central nervous system. Moreover Sherrington, Bolton, and others had already put forward the hierarchical theory, and had demonstrated that the *general* structure appeared much the same throughout the nervous system for any given individual, but that there were also minor *local* variations. As for genetics, Mendel's multifactorial theory of inheritance was adopted almost from the outset. Nor have later factorists (at any rate in this country) shown the smallest inclination to " move from this universe ". When environmental influences have been kept approximately constant by statistical or experimental devices, the ' vectors ' or ' factors ' are in fact regularly identified with hypothetical genetic tendencies influencing the structure and functioning of the brain. And what more can Dr. McLeish desire ?

The next chapter, entitled ' Heredity and Environment ', somewhat inconsistently proceeds to criticize the importance of genetic influences. " The problem of the relative importance of heredity and environment in the determination of individual intelligence ", we are told, " is now seen to be *not* a scientific question." The views of those who maintain that individual differences are largely of genetic origin " are based on social-political bias ", " middle-class prejudice masquerading as science ". As a result, it is said, " working-class children " are badly handicapped, and thousands of " working-class students deprived of a university education . . . quite arbitrarily and unjustly ". Once again the chief object of his attack is the work of Burt, Ballard, and their colleagues under the London County Council.

Dr. McLeish starts with a tirade against the alleged ' constancy of the intelligence quotient '—' the immutable adamantine IQ of Ballard '. Burt, it is said, " *now* admits that intelligence as tested may fluctuate ". But Dr. McLeish does *not* add that he has always done so, and that 40 years ago Burt's original memorandum on the Binet tests contained a section, with illustrative cases, emphasizing ' changes in the mental ratio '.[1] He then reproduces Burt's earlier table of the average assessments of the intelligence of parents and children from different occupational classes (intended primarily to show how the children regress towards the mean). He omits the standard deviations, and ignores Burt's remark that " the group differences shown in the table may be a little misleading unless we bear in mind the wide range of individual ability within each class," as well as the emphasis placed on the *low* correlation between class and intelligence. " We must," he continues, " assume for Sir Cyril's case that the children of the casual labourer have the same opportunity as the professional man's child of hearing and using such words as ' munificent ', ' malevolent ', etc.,—to take words in one of the subtests . . . These items are typical of the tests used to demonstrate differences in intelligence between social classes."

True to his principles, Dr. McLeish gives no books or articles to enable his readers to check his statement. The table he quotes appears in a number of different publications, e.g. in the paper already cited, in Conway's articles on ' Class Differences ' in this journal, and in Burt's later article on ' Ability and Income ' in the *British Journal of Educational Psychology* (XIII, p. 84—surely not such an " out-of-the-way periodical "). None of these supports Dr. McLeish's interpretation of their views; and it is difficult to imagine how he came by them. The survey was carried out by a team of which I myself was a member. For the preliminary screening we used a group test with verbal and non-verbal subtests; the individual testing was carried out with the Stanford-Binet scale; and " the crude test-results were always submitted to the children's teachers and checked in other ways". The test that Dr. McLeish describes (not quite correctly) seems to have been one of the group tests used for *adult* entrants to the Arts and Science Faculties of the University:

---

[1] *Mental and Scholastic Tests*, pp. 151f. In an earlier report (1917) he had already deprecated the use of the ' intelligence quotient ' as a unit except for popular exposition, and recommended using the standard deviation of the age groups, which demonstrably shows greater constancy, though even that, we are told, is far from perfect. No page–references are given to support Dr. McLeish's interpretation of Ballard's statements.

it would of course have been ludicrous to apply it to " the children of the casual labourer " in the way Dr. McLeish describes and critizes—or indeed to any children. What we sought to discover in that particular survey was—how many children, of high ability as demonstrated by tests and other methods, nevertheless failed as a result of environmental handicaps to win a place in secondary schools of the grammar type (for which the entrance examination in those days was mainly scholastic, i.e. English and Arithmetic) and a fortiori failed to obtain the places in a University to which their abilities entitled them. The estimated number was 40 per cent, chiefly from the socalled ' manual ' classes. This is hardly a conclusion that would imply a " social-political bias " of the type Dr. McLeish imputes to us. Let me add that the method of standardized testing, and the supplementary devices for estimating children's potential capacities introduced by the London County Council, must have saved hundreds of handicapped children from being certified as mentally defective, rescued a good many from delinquency, and enabled (as the records showed) large numbers from poverty-stricken homes to gain places in grammar schools and eventually to secure entrance to the University.

The next chapter deals with ' the distribution of intelligence '. Here Dr. McLeish's habit of assuring us that those whom he criticizes have assumed or stated almost exactly the opposite to what they really did assume or state produces still further surprises. " One of the most basic assumptions of the constructors of intelligence tests ", he tells us, " is that intelligence is distributed in the population in accordance with the law of normal probability . . . Without making this assumption it is impossible to calculate intelligence quotients . . . Yet there is no known way by which this assumption can be verified or disproved." Intelligence quotients are calculated by dividing a child's mental age by his chronological, and then multiplying the ' quotient ' by 100 to convert it to a percentage: why it is impossible to make such calculations without making assumptions about their distribution is never explained. The chi-squared test provides a well-known way of " verifying or disproving " whether or not the distributions so obtained agree with the assumption of normality. From the days of the earliest surveys (1917) this statistical criterion has been regularly applied; and, wherever the sample has been large enough to furnish the requisite data, it has always turned out that individual differences were *not* distributed normally, but (as we might expect) in accordance with a curve of Type IV. When dealing with extreme deviations—the numbers of the mentally deficient or the highly gifted—the assumption of normality leads to very erroneous estimates. Within a more limited range the normal distribution supplies fairly reasonable approximations. But the reason for assuming normality is not, as Dr. McLeish suggests, because mental testers adopted it " as an axiom ", but simply because the frequencies for the normal curve have been tabulated, whereas the frequencies for other types have to be calculated afresh, and the calculation is both arduous and lengthy.

Dr. McLeish then attacks the various assumptions made in the paper on ' The Multifactorial Theory of Inheritance ' by Burt and myself (this *Journal*, IX, pp. 95f.), in which we endeavoured to apply the method devised by Fisher for studying the inheritance of physical characteristics (e.g. stature) to the inheritance of differences in mental abilities. This method, he complains, " rests on a now outmoded genetic theory. It is assumed that, if we have a large number of genes which carry the quality of ' intelligence ', and these are distributed in a random fashion in a population which mates at random . . . this will result in each successive generation having ' intelligences ' which cluster round the average with a symmetrical distribution—in other words, the Gaussian bell-shaped curve . . . It is quite clear ", he adds, " that no human group conforms to these abstract conditions".

This is hardly a fair account of our procedure. " The theoretical evidence ", we argued, " suggests that differences in intelligence are the effect *partly* of multifactorial and *partly* of unifactorial inheritance: on this basis the distribution expected would be a bell-shaped curve, more or less skewed, i.e. type IV." Our actual demonstration, however, turned not on a comparison of frequency-curves but rather on an analysis of variance and a comparison of correlations. Fisher's original method, which is by no means 'outmoded', was brought up to date in accordance with the 'modifications' suggested by Mather and others. It is true in the first section of our proof we began by assuming 'random mating'. But in the following section additional complications were introduced into the formulae to allow for 'assortative mating' (loc. cit., pp. 113f.). The systematic comparison of the theoretical values with the empirical data seemed fully to confirm the theory. If Dr. McLeish wishes to disprove it, he should surely undertake a similar analysis of his own, and not airily dismiss them on the basis of a misleading caricature. Naturally, if a writer does not trouble to learn the statistical techniques involved, the arguments which are based on them must inevitably seem to "cloud his understanding". But this does not excuse the misrepresentations; and Dr. McLeish's protests surely carry the demand for 'simplicity in science' a little too far.

Dr. McLeish ends this part of his book with a chapter on 'the disintegration of intelligence test theory'. He starts with the assertion "that the metaphysical nature of the concept of innate intelligence has been explicitly recognized by Sir Cyril Burt", and, we gather, by several of his co-workers and followers. Vernon's 'Intelligence A', "which can be identified with Burt's 'innate intelligence'," is also "removed from the plane of scientific discourse to the realm of metaphysics." Once again we are never told where this "explicit recognition" is to be found. Burt has a chapter on 'metaphysical' issues in *The Factors of the Mind*, and denies the very proposition that Dr. McLeish attributes to him. "Intelligence", he says, "I regard as specifying certain individual differences *in the structure of the central nervous system*—differences whose concrete nature could be described in histological terms"; and it is on this ground that he assumed that such differences, like other structural differences, are "largely determined by innate tendencies." This may be hypothetical, but it is certainly plausible, and in no sense 'metaphysical'.

"The intelligent layman" to whom the book is primarily addressed will by now have gathered that the honesty of the whole tribe of mental testers is open to question. Spearman is accused of "an elaborate statistical cooking of the results of tests to conform with his preconceived notions". Burt's methods of estimating the intelligence of individual children are said to have been "of a really shocking character" (though most of it was done before Dr. McLeish himself left school, so that he can have no first-hand knowledge, and no one else has made such a complaint). The rest of us are accused of "smuggling in idealistic notions", presumably to support our alleged middle class prejudices (though a good many of us happen to be of working class origin ourselves).

If I appear to have discussed Dr. McLeish's criticisms of our work under the London County Council at a somewhat excessive length, that is because his assertions have already been taken at their face value, not only by educationists and lecturers at training colleges, but by many members of the general public who suppose that here at last the fallacies of mental testing, school streaming, and selection at 11 plus have once and for all been exposed by a fellow psychologist. It would seem essential therefore to warn readers not to accept any statement made by Dr. McLeish on this subject without first consulting the official reports and published articles to which he is apparently referring, or (if they cannot trace the references) to ask Dr. McLeish to supply chapter and verse.

The remainder of the work deals with topics on which a statistical psychologist like myself is scarcely competent to speak—'the science and art of education', 'a model for man' (with 'five prototypes'), and (much more briefly) existentialism, psychoanalysis, and the views of Marx, Engels, and Lenin. The book ends with a chapter entitled (like the book itself) 'the science of behaviour' and inevitably another tilt at the iniquities of intelligence testing. And once again throughout these chapters no references are given to enable the student to follow up the various subjects or the inquisitive critic to verify the accuracy of Dr. McLeish's interpretations.

M. HOWARD.

Social Psychology Through Experiment. Edited by G. HUMPHREY and M. ARGYLE. London: Methuen. 1962. Pp. iv + 208. 21*s.*

In their preface to this book the editors explain that, since in the teaching of any scientific subject laboratory work should always form an important part of the course, they have attempted to supply a "handbook and guide to practical class work in social psychology". Seven main topics have been selected; and for each a specialist has been invited to describe three experiments suitable for laboratory exercises. In every case the descriptions are prefaced by a discussion of the prevailing theories and of the relevant facts in the particular field concerned, and are followed by a short bibliography. Some of the experiments—e.g., that on 'friendship choices at school'—have to be carried out in local schools; and several of the others—e.g., those on the effects of competition, cooperation, and the interpretation of personal interviews—should preferably be carried out with children, since an ordinary university class is not likely to include a sufficient number or a sufficient variety of personalities to make the results really instructive.

The editors are anxious that none of their readers should be 'startled' by the novelty of their project; and accordingly, they provide two introductions—one 'for humanists' and the other 'for psychologists'. However, the proposal is by no means so new as they suppose. In training colleges, where practical courses in psychology have long been provided, attempts have commonly been made to illustrate the social aspects of the subject; and experiments similar to many of those in the present volume were introduced long ago, with adult students as subjects as well as school children. The purpose was much the same—to elucidate the influence of groups on communication, cooperation, efficiency of work, accuracy of judgement, and the mutual interactions of personalities.

The authors' choice of problems and their general line of discussion are likely to excite a good many questions in the minds of experienced teachers of psychology, particularly those who are insistent on the adequate design of experiments and a scrupulous selection of adequate statistical procedures. Some of the discussions are also open to criticism on the ground of terminology. The opening chapter, for example, deals with experiments on 'social perception'. This phrase, we are told, "stands for studies attempting to find regular and predictable relations between man's general orientation towards his environment and the multitude of social factors determining it". The first of the experiments under this head is intended to demonstrate the way in which the recognition of objects is facilitated by the possession of distinctive names. Such knowledge, however, is not in itself a 'social factor': (e.g., I myself learnt to recognize the chief constellations in the sky, and still recognize various types of physiognomy, by inventing *my own* names for them). The second experiment requires a number of bilingual people to record on tape statements first in one language (e.g., English), then in another (e.g., French). The listeners, who do not know that each person speaks twice, are then asked to evaluate the personalities; and it is expected that their stereotyped notions of the characteristics of the different nations will tend to affect their judgement: in this case the practical details of the classroom procedure are very vaguely described. In the third experiment students are given one or more traits (e.g., 'intelligence') and asked to indicate what else they would infer from the possession of these traits by marking a list of 59 other traits (e.g., 'clever', 'intellectual', 'efficient', 'friendly', etc.). Whether 'social' or not, experiments such as these have nothing to do with 'perception'.

Many of the experiments must plainly consume a good deal of time; and, considering the uncertain results likely to be obtained, they will often convey little information to the students who take part. Hence a good many teachers will probably wonder whether after all the time required might not be better spent in other ways. With beginners an adequately planned investigations is far easier to explain than to carry out, and the well established results of such investigations can be summarized in a few minutes, whereas it may take half-an-hour to explain why, owing to uncontrolled variables, the students themselves obtained a misleading result. The object of one of the experiments, for example, is to determine whether a large group is more effective at solving problems than small groups; four groups—two containing three students and two six—are required to solve fifteen problems by the method of ' Twenty Questions '; four rooms, four experiments with stop watches, and four tape recorders are needed. To a statistical investigator the instructions given and the discussions that follow will appear rather naïve. And how precisely the data obtained are to be analysed is by no means obvious. It is tentatively suggested that comparisons might be made either by the chi-squared test or by the Wilcoxon signed rank test (which is not described). As it happens, this is a class experiment I myself have been able to watch: to my mind it furnished an admirable example of how such a research should *not* be planned or evaluated.

But it is always easy to pick holes in a pioneer project. My criticisms are meant to raise questions rather than to formulate objections. Whether or not the book can be recommended for its original purpose may perhaps be doubted; but it will certainly provide every teacher of the subject with food for reflection and material for study. J. C. SCOTT

British Journal of Industrial Relations. Vo. I, No. 1. Edited by B. C. ROBERTS. London: The London School of Economics and Political Science. 1963. Pp. 136. 15*s*. (42*s*. per volume of three parts; U.S.A. $6.50).

During the years that have elapsed since the close of the last war there has been a growing and widespread interest, academic as well as personal and political, in the problems of industrial relations. The topic is already recognized as an appropriate field both for teaching and for research; and at least one British university now has a department devoted to the subject. With the rapidly increasing mass of valuable material awaiting publication it has for some time been clear that a new periodical was badly needed.

" The field to be covered ", we are told, " will not be limited by any narrow definition of the phrase ' industrial relation ' ". Historical studies, as well as papers reporting the results of systematic inquiries or research on contemporary questions, will be accepted. Although the general presentation will be academic, it is hoped that the material published will appeal to a much wider field of readers.

The problems outlined in the editorial note which introduces this journal include many to which statistical psychologists might well make valuable contributions; and the first issue contains several articles which will undoubtedly interest many of our readers— ' Communications in Trade Unions ', ' Political Independence in British Trade Unions ', 'A Study of Soviet Wages ', and ' Social and Labour Policy in the European Economic Community '. There is a detailed chronicle of relevant events, month by month, throughout 1962, and half-a-dozen illuminating book reviews. The journal is admirably edited, printed, and produced. Psychologists, and indeed all who are concerned with the social sciences, will welcome its appearance. CYRIL BURT.

INDEX TO VOLUMES I TO XIV

An index to the first fourteen volumes of the *British Journal of Statistical Psychology*, classified according to authors and subjects, is now available. As proprietors of the journal the British Psychological Society wish to take this opportunity of expressing their gratitude to Mrs. Beatrice Warde to whose generosity and expert assistance the compilation and printing of the index is due.

OCCUPATIONAL PSYCHOLOGY

Editor : Alec Rodger

JULY, 1962 — CONTENTS — VOLUME 36, No. 3

Operational Research, Psychological Information and Managerial Decisions — By R. A. M. Gregson

Extra-mural Factors influencing Behaviour at Work — By Sylvia Shimmin

Youth Employment Service Interviews: Part 1, A Criterion Problem — By P. Cavanagh, R. I. Drake and K. F. Taylor

The Relation Between Method of Wage Payment and Incidence of Psychosomatic Disorders — By Leonard J. Fletcher, Jr. and J. Richard Simon

Personality Ratings by Faculty as Predictors of Executive Success for Graduate Business School Alumni — By Albert Porter

Some Hidden Determinants of Inspectors' Judgments — By P. Cavanagh and Alec Rodger

The Design of Information — By R. Conrad

Safety Harnesses for Road Vehicles — By J. P. Dennis and H. C. W. Stockbridge, with the assistance of C. Q. Large

Book Reviews

Other Books Received

The H.M. Vernon Prize

Annual Subscription **40** *shillings*

National Institute of Industrial Psychology, 14 Welbeck Street, London, W.1

9 DEC 1963

BRITISH JOURNAL OF STATISTICAL PSYCHOLOGY

Vol. XVI Part 1 May 1963

CONTENTS

		Page
CHRISTENSEN, PAUL R. AND GUILFORD, J. P.	An Experimental Study of Verbal Fluency Factors	1
YOUNG, MICHAEL AND GIBSON, JOHN	In Search of an Explanation of Social Mobility	27
HALLWORTH, H. J. AND WAITE, G.	A Factorial Study of Value Judgements among Adolescent Girls	37
NOTES AND CORRESPONDENCE		
	The Psychology of Value	47
	Ability and Attainment	105
CRITICAL NOTICE	The Use of Electronic Computers in Psychological Research	118
BOOK REVIEWS		126

Vol. XVI Part 2 November, 1963

THE BRITISH JOURNAL OF STATISTICAL PSYCHOLOGY

EDITED BY
R. J. AUDLEY

WITH THE ASSISTANCE OF
CYRIL BURT

AND THE FOLLOWING EDITORIAL BOARD

C. BANKS
M. S. BARTLETT
G. H. BOWER
D. E. BROADBENT
J. BROWN
V. R. CANE
D. R. COX
L. J. CRONBACH
C. I. HOWARTH
A. R. JONCKHEERE
D. N. LAWLEY
R. D. LUCE
S. MESSICK
E. A. PEEL
L. S. PENROSE
J. FRASER ROBERTS
W. STEVENSON
A. STUART
P. E. VERNON

Printed and Published by
TAYLOR & FRANCIS LTD.
RED LION COURT, FLEET STREET, LONDON, E.C.4

Price 20*s. per part* (*U.S.A.* $3.50) *Subscription* 30*s.* 6*d. per volume* (*U.S.A.* $4.50)

PUBLICATIONS OF THE BRITISH PSYCHOLOGICAL SOCIETY

The British Journal of Statistical Psychology is issued by the British Psychological Society. The subscription price per volume is 30*s*. net for Members of the Society and 30*s*. 6*d*. (post free) for non-members. The subscription price in the U.S.A. is $4.50 (post free). Members of the Society should send their subscriptions to THE SECRETARY, British Psychological Society, Tavistock House South, Tavistock Square, London, W.C.1: non-members to Messrs. TAYLOR & FRANCIS, LTD., 18 Red Lion Court, Fleet Street, London, E.C.4. Until further notice the Journal will be issued in two parts each year, about May and November.

Papers for publication and books for review should be sent to the Editor, R. J. Audley, Department of Psychology, University College London, Gower Street, W.C.1. Authors should retain copies of their manuscripts as neither the Society nor its Editors can undertake any responsibility for loss or delay in the case of unsolicited contributions. Authors will receive 50 copies of their articles free; extra copies may be purchased provided the order is given when the proofs are returned. Contributors should indicate whether they wish their reprints to be bound in covers. The printers will state current prices when sending out proofs.

The Society also issues quarterly *The British Journal of Psychology* and *The British Journal of Medical Psychology*, at a subscription rate of 60*s*. per annum for either journal, and *The British Journal of Social and Clinical Psychology* three times a year at a subscription rate of 50*s*. per annum. Subscriptions for these journals should be sent to the Cambridge University Press, Bentley House, Euston Road, London, N.W.1. The Society publishes, jointly with the Association of Teachers in Colleges and Departments of Education, *The British Journal of Educational Psychology* three times a year: for this the subscription, 30*s*. per annum, should be sent to Methuen & Co. Ltd., 36 Essex Street, London, W.C.2. Members of the Society receive the foregoing journals on special terms; enquiries should be addressed to the Secretary, British Psychological Society, Tavistock House South, Tavistock Square, London, W.C.1.

Papers for publication in *The British Journal of Psychology* should be sent to Dr. Boris Semeonoff, Department of Psychology, The University, Edinburgh 8, those for publication in *The British Journal of Medical Psychology* to Dr. Thomas Freeman, Lansdowne Clinic, 4 Royal Crescent, Kelvingrove, Glasgow, C.3; those for publication in *The British Journal of Social and Clinical Psychology* to Mr. Michael Argyle (Social Psychology Editor), Institute of Experimental Psychology, 1 South Parks Road, Oxford, or to Dr. Jack Tizard (Clinical Psychology Editor), Maudsley Hospital, Denmark Hill, London, S.E.5; and those for publication in *The British Journal of Educational Psychology* to Mr. L. B. Birch, University of Sheffield, Institute of Education, Sheffield 10.

PREPARATION OF PAPERS

Contributors are asked to send their papers in a form which is ready for submission to the printer: that is to say, the arrangement of the material should follow that observed in previous publications of this *Journal*. Each article should be headed either with a series of numbered headings in italic, corresponding with the cross-headings of the several sections, or with a brief abstract: (the latter procedure is preferred). Each section should have a short cross-heading, in capitals; and the whole should conclude with a summary embodying the main conclusions reached. Special attention should be paid to such details as correct spelling, grammar, capitalization, numbering, etc. (particularly in the case of tables and references). In preparing manuscripts and in correcting proofs authors should conform, so far as possible, with the 'Recommendations to Authors' set out in *The Printing of Mathematics* by T. W. Chaundy, P. R. Barrett, and Charles Batey (Oxford University Press, 1954, ch. II, pp. 21–73), or to those given in the paper on 'Setting Mathematics' (Arthur Phillips, *Monotype Recorder*, XL, iv, 1956).

Owing to the present conditions for printing, the cost of setting up tables and algebraic formulae which require hand composition is very much greater than the cost of machine composition. Contributors are therefore advised to arrange their material so that it can be printed as economically as possible. Often numerical results and algebraic formulae can, with a little simplification, be readily adapted for purposes of machine composition. Manuscripts involving many tables and numerous equations can as a rule only be accepted if the authors are willing to pay for the additional cost. Contributors should note that the *Journal* is intended for the general psychological reader, not solely for those who specialize in statistics. Unusual technicalities should be either avoided or carefully explained, e.g. by reference to previous discussions of the same or kindred problems in this *Journal*.

In future authors will receive one complete slip-proof of their papers, but no page proof. They will be expected to pay the cost of all corrections for which they themselves are responsible.

All manuscripts should be submitted in duplicate.

Vol. XVI
Part 2
The British Journal of Statistical Psychology
November
1963

THEORY OF GENERALIZABILITY: A LIBERALIZATION OF RELIABILITY THEORY[†]

By Lee J. Cronbach, Nageswari Rajaratnam[‡] and Goldine C. Gleser
University of Illinois

" Reliability theory " is reinterpreted as a theory regarding the adequacy with which one can generalize from one observation to a universe of observations. If the observation is randomly sampled from the universe—whether or not the universe consists of equivalent observations—the intraclass correlation provides an approximate lower bound to the expected value of the desired coefficient of generalizability.

The theory and procedures by which the psychologist judges measuring instruments developed largely out of Spearman's research on mental tests at the turn of the century. The classical theory, which centres on reliability and validity (identified with interform correlations and test-criterion correlations, respectively), has been very fully elaborated [13]. In recent years, many writers have re-examined and extended the definition and interpretation of validity, motivated chiefly by dissatisfaction with the classical supposition of a single observable criterion against which a test is to be judged. There has been no similarly fundamental reconsideration of reliability theory, though parts of the theory have been restated from time to time.

The heart of classical reliability theory is the concept of ' parallel ' measures. Just as classical validity theory postulates a criterion, so classical reliability theory postulates two or more measures equivalent in content, in mean, in variance, and in intercorrelations. In internal-consistency analysis, it is further supposed that half-tests or even separate items satisfy these (or slightly weaker) equivalence assumptions. For standardized tests, parallel forms can be constructed that very nearly satisfy the equivalence conditions; but test items are usually heterogeneous, and half-tests may likewise fall short of equivalence. For many other types of measurement, parallel observations are hard to attain and sometimes the very concept of parallel observations is obscure. We often wish to evaluate the dependability of ratings, but one rater is rarely identical to the next in what he attends to, in the variance of the ratings he gives, or in his correlations with other raters. An investigator observing behaviour in the field encounters a similar difficulty; different situations in which the subject is observed are rarely equivalent.

[†] Prepared under grant M-1839 from the National Institute of Mental Health. Victor McGee and John Hunter have contributed to this report; we acknowledge also the contributions of several consultants and correspondents, particularly F. M. Lord and H. E. Brogden. Our theory has much in common with a ' schematization ' presented by O. K. Buros to the American Educational Research Association, 1963; we have been asissted by draft versions of his work.

[‡] Now at the University of Minnesota.

This paper offers a model that does not presuppose equivalence of observations. The formulation is sufficiently general to encompass nearly all earlier work, and therefore clarifies the application and interpretation of traditional reliability formulas. We find that well-known formulas serve satisfactorily under the new model, save when measures are very far from equivalent.

I. ASSUMPTIONS UNDERLYING PREVIOUS DEVELOPMENTS

Classical Theory. As notes on the history of reliability theory have recently been published by Burt [4] and Tryon [37], we shall review only the evolution of assumptions. Spearman is the father of the classical reliability theory in psychology. Because correlations between measures of different attributes are attenuated (reduced) by errors of observation, Spearman needed to correct the observed intertest correlations in order to determine how closely the postulated underlying abilities were correlated. In his 1904 paper [33] he proposed to correct the observed $\rho_{x_1y_1}$ by dividing it by $\sqrt{(\rho_{x_1x_2} \cdot \rho_{y_1y_2})}$. In this first formulation, it was demanded only that x_1 and x_2 be observed independently and that in some undefined sense they estimate the same thing; likewise for y_1 and y_2. Formal equivalence conditions did not appear. Indeed, Spearman considered that x_1 might be more (or less) accurate than x_2, and allowed measures of the same quality to differ in content. One example given was a set of three alternative tests of 'discrimination' that required, respectively, judgments of pitch, brightness, and weight. At this time, reliability estimates were obtained solely to make a correction for attenuation; it was only in a footnote that Spearman suggested using $\rho_{x_1x_2}$ to describe the accuracy of observations.

Spearman's 1910 paper [34] was more concerned with reliability. He considered a series of measures of the same attribute, the score on each test having been generated by combining a random error with the same 'true score'. It was assumed that the errors had a mean of zero, had uniform variances, were independent, and were independent of true score. Under these conditions all tests have equal means and variances, and the parallel-test correlations are (i) equal, (ii) equal to the ratio of true-score variance to observed variance, and (iii) equal to the squared correlation of observed score with true score. Furthermore, from these assumptions it is possible to derive the formula for the increase in reliability of a composite as the number of measures increases, and the correction for attenuation [13, Ch. 2].

Whereas Spearman based his development on the true-score-plus-error assumption, Brown [2] began by defining parallel tests. His approach leads to the same reliability theory as Spearman's [13, Ch. 3] What is postulated by one is derived by the other; such small logical distinctions between the theories as were once matters for contention [see 37] no longer seem important. Whichever starting place is chosen, the true score turns out to be the limit of the mean observed score as the number of tests becomes indefinitely large.

In 1910, Spearman required that the measures of x be so alike that any difference among the several scores for the same person 'may be regarded as quite accidental' (p. 274). Gulliksen states this requirement more elaborately (p. 14): "In addition to satisfying these objective and quantitative criteria (equal means, variances, and intercorrelations), parallel tests should also be similar with respect to test content, item types, instructions to students, etc." This restriction on content is nowhere embodied in the mathematical model of classical theory; nothing in the classical mathematical assumptions prohibits each test from having some unique psychological content. Thus a series of compositions on diverse topics, used to appraise writing ability, would not have uniform content; it might nonetheless conform to the mathematical model.

Internal-consistency formulas. Spearman [34] developed a formula for estimating the correlation between two parallel series of measures, when only one series of measures has actually been made; Brown independently arrived at the same result, subsequently known as the 'Spearman-Brown' formula. Spearman proposed that a series of measures of verbal memory, or a series of ratings of a child by a teacher, could be split into two separate parts. The correlation between the scores for the parts, corrected by his formula, would estimate the correlation of the whole series with another equivalent whole series. Other workers applied this strategy to a single test rather than a series of tests; dividing the test into odd and even items provided two sets of scores to which Spearman's formula could be applied. The rationale required the half-series or half-tests, observed and hypothetical, to be experimentally independent and equivalent.

Another split-half formula takes this form:

$$\text{reliability} = 1 - \frac{V_{A-B}}{V_T} = 2\left[1 - \frac{V_A + V_B}{V_T}\right] \tag{1}$$

where A and B are the halves of the test T, and V symbolizes variance. This formula apparently was first suggested by Flanagan, who published no derivation. Rulon [31] advanced an argument for regarding the variance of errors of measurement as equal to V_{A-B}; this leads to the foregoing formula for the ratio of true-score variance to observed variance. Rulon assumes that the scores A and B have equal covariances with the true score and that errors of measurement in the two halves are uncorrelated with each other and with the true score. V_A and V_B are not assumed to be equal, but Rulon does appear to identify reliability with the correlation between equivalent whole tests.

Kuder and Richardson [24] developed various formulas using item statistics of a single test to estimate the correlation between equivalent tests. The most important of their formulas are KR20 and KR21. KR20 yields the intertest correlation if the product-moment correlation matrix of all $2n_i$ items is of unit rank (no group factors) and if items i, j in one test correspond respectively to items i', j' in the second test in such a way for every pair i, j that $r_{ij} = r_{ij'} = r_{i'j} = r_{i'j'}$. An assumption of uniform item variance is added to derive KR21 from KR20. In these formulas, items are assumed to be dichotomous (scored 1 or 0). With dichotomous items, however, the unit-rank requirement cannot be strictly satisfied unless all items have the same mean and hence the same variance. Thus the original derivations of *both* KR21 and KR20 assumed equal item means, equal item variances, and single-factoredness of items. Subsequent writers [cf. 8, 19] have replaced KR20 with a general formula not restricted to dichotomous items, one of whose form is:

$$\text{reliability} = \frac{n_i}{n_i - 1}\left(1 - \frac{\Sigma V_i}{V_T}\right) \tag{2}$$

When n_i—the number of separately scored parts—is 2, this reduces to (1).

Jackson and Ferguson [21] showed that (2) gives the correlation of the test in hand with an equivalent test, assuming only that the mean within-test item covariance is the same for both tests and is equal to the mean of the cross-test item covariances (that is, $\bar{C}_{ij} = \bar{C}_{i'j'} = \bar{C}_{ij'}$). This condition seems weaker than those used by Kuder and Richardson, but it actually departs very little from the unit-rank assumption. Cureton [11], in pointing out the restrictive nature of the Jackson-Ferguson assumption, went too far in stating that the assumption always implies unit rank; hypothetically, tests with several factors may satisfy the condition if their items have certain symmetric patterns of factor loadings. Such exceptions appear to be possible, however, only where the number of equivalent tests is finite; even then, such exceptional patterns are unlikely to occur in practice.

H. F. Kaiser (personal communication), interpreting the squared correlation of test score with universe score (defined as we define it below) as a reliability coefficient, shows that formula (2) results from an assumption somewhat weaker than that of Jackson and Ferguson. If i and j are items within the test, and p and q other items in the universe, he assumes that $\bar{C}_{ip}{}^2 = \bar{C}_{ij}\bar{C}_{pq}$. Just how this constrains the factorial structure of the universe of items is not known.

Guttman [14] shows that (1) and (2) are lower bounds, in the population, to the correlation between measures obtained by experimentally independent applications of the same test. The several measures are thus assumed to be equivalent, but no assumption about equivalence of items or half-tests is made.

To summarize: The classical approach defines reliability as the correlation between equivalent measures. This correlation can be estimated from part-scores within one measure. In the earliest derivations of internal-consistency formulas the parts (half-tests or items) as well as the whole tests were assumed to be equivalent; later work showed that the same formulas result under weaker assumptions about the part-scores.

Intraclass correlation. The first proposal to use an intraclass correlation as a reliability coefficient apparently came from Cureton [10, p. 24], but the main source of reliability theory based on variance ratios was the University of London. Jackson [20], working with Neyman and Johnson, used their statistical concepts to appraise precision of measurement. After Johnson's return to Minnesota, his student Hoyt [19] extended the concepts to internal-consistency analysis. At the same time, Burt, consulting with Fisher and other members

THE BRITISH JOURNAL OF STATISTICAL PSYCHOLOGY

EDITED BY
R. J. AUDLEY

WITH THE ASSISTANCE OF
CYRIL BURT

AND THE FOLLOWING EDITORIAL BOARD

C. BANKS
M. S. BARTLETT
G. H. BOWER
D. E. BROADBENT
J. BROWN
V. R. CANE
D. R. COX
L. J. CRONBACH
C. I. HOWARTH
A. R. JONCKHEERE
D. N. LAWLEY
R. D. LUCE
S. MESSICK
E. A. PEEL
L. S. PENROSE
J. FRASER ROBERTS
W. STEVENSON
A. STUART
P. E. VERNON

VOLUME XVI

Printed and Published by
TAYLOR & FRANCIS LTD.
RED LION COURT, FLEET STREET, LONDON, E.C.4

of the statistics department at London, elaborated a similar theory in unpublished laboratory notes [cf. 3, 6]. Numerous papers on reliability subsequently made use of the analysis of variance or the closely related intraclass correlation; a partial list includes Hoyt [19], Burt [3, 4], Jackson and Ferguson [21], Alexander [1], and Ebel [12]. The analysis of variance (anova) can be extended to complex designs involving multiple sources of error (e.g., test forms, occasions, and scorers considered simultaneously). The most complete published treatments are those of Lindquist [25] and Medley and Mitzel [28, pp. 309–317]. Rather than trace detailed differences between these papers, we shall summarize the line of thought that emerged from them, limiting ourselves to singly-classified sources of error.

Let X_{pi} be the observed score of person p on test i (or given by rater i). Let M_p be the universe (' true ') score of person p (the mean of X_{pi} over all i in the universe), M_i the mean score on i over all persons in the population, and M the mean of the M_p or the M_i (the grand mean). Whereas Spearman began with an assumption of the form

$$X_{pi} = M_p + e_{pi}\,, \tag{3}$$

Jackson and successors regarded X_{pi} as the sum of four components:

$$X_{pi} = M + (M_p - M) + (M_i - M) + e_{pi}\,. \tag{4}$$

The form of this identity corresponds to the additive model in anova where the error terms e_{pi} are usually considered to be independently distributed random variables with zero means and equal variances. The main effects $M_p - M$ and $M_i - M$ may be regarded either as fixed unknown constants or as random variables that are independent of each other and of the error terms.

Most treatments of reliability based on (4) define reliability as the ratio of true score variance V_{M_p} to observed-score variance in the population; this ratio is an intraclass correlation. There is a certain ambiguity in the phrase 'observed-score variance'. If all persons take the same test i or are rated by the same rater i the score variance is $(1/n_p)\sum\limits_p (X_{pi} - (1/n_p)\sum\limits_p X_{pi})^2$. If each person is observed under n_i conditions, and the resulting scores are thrown into a single distribution, the variance of these observed scores is $(1/n_p n_i)\sum\limits_p\sum\limits_i (X_{pi} - (1/n_p n_i)\sum\limits_p\sum\limits_i X_{pi})^2$. The limit of the first variance—i.e., $\lim(n_p \to \infty)\ (1/n_p)\sum\limits_p (X_{pi} - M_i)^2$—we shall label V_i. If we think of scores as arranged in a matrix with persons as rows and conditions as columns, this is the variance of scores within a particular column, for an infinite number of persons. The limit of the second variance—$\lim(n_p, n_i \to \infty)\ (1/n_p n_i)\sum\limits_p\sum\limits_i (X_{pi} - M)^2$—we shall label V_X. It is the variance of all the scores in a matrix with an infinite number of conditions and persons.

These variances lead us to two intraclass correlations, V_{M_p}/EV_i (E standing for 'expected value of') and V_{M_p}/V_X; these we shall label α and $\rho^2{}_{MX}$. (Formerly α has been used to designate the reliability estimate obtained by formula (2), which applies to composites of n_i items or part-scores; in this paper we shall designate that result $\hat{\alpha}_{(n_i)}$).

The α coefficient is estimated from an array of scores on n_i tests for n_p persons.* The customary two-way analysis of variance gives a between-persons mean square MS_p and a residual mean square MS_r. These mean squares are inserted in the following estimation formulas:

$$\left.\begin{aligned} \hat{V}_{M_p} &= \frac{1}{n_i}(MS_p - MS_r) \\ \hat{V}_e &= MS_r \\ \hat{E}V_i &= V_{M_p} + V_e = \frac{1}{n_i}[MS_p + (n_i - 1)MS_r] \end{aligned}\right\} \tag{5}$$

Consequently,

$$\hat{\alpha}_{(1)} = \hat{V}_{Mp}/\hat{E}V_i = (MS_p - MS_r)/(MS_p + (n_i - 1)MS_r) \tag{6}$$

The subscript in $\hat{\alpha}_{(1)}$ indicates that this is the reliability of a single score, not a composite. It is sometimes said [see 3] that derivations of this nature avoid the classical equivalence assumptions. But the variance estimates in (5) are traditionally reached by means of models that assume that the e_{pi} are random variables with common variance V_e. This forces the V_i to be equal and the correlations between tests to be equal. Hence the derivation of (6) from (4) assumes equivalence of conditions except for allowing test means M_i to vary.

Formula (6) can be employed whenever there are at least two test scores for every person in the sample. Where persons take just one test, an internal-consistency formula can be applied, in which the n_i items or subtests are used as separate observations. Then for the total test score the estimated true variance $\hat{V}_{Mp} = n_i(MS_p - MS_r)$ and the expected variance of observed scores is $n_i MS_p$; consequently,

$$\hat{\alpha}_{(n_i)} = (MS_p - MS_r)/MS_p \tag{2'}$$

This formula is numbered (2′) because it yields the same result as eqn. (2) (the generalized KR20).

If reliability is defined as V_{M_p}/V_X, the variances are estimated by eqns. identical to (5) except that MS_r is replaced by the within-persons mean square MS_{wp}. This may be obtained directly from a one-way analysis of variance, or by combining the residual and the between-persons sums of squares from a two-way analysis and dividing by $n_p(n_i - 1)$. The reliability of a single score is then

$$\hat{\rho}^2_{MX_{(1)}} = \hat{V}_{M_p}/\hat{V}_X = (MS_p - MS_{wp})/(MS_p + (n_i - 1)MS_{wp}) \tag{7}$$

and that for a composite is

$$\hat{\rho}^2_{MX_{(ni)}} = (MS_p - MS_{wp})/MS_p \tag{8}$$

Equation (3), when applied to a set of dichotomous items, approaches KR21 when the number of persons n_p is large. Kaitz [22], who noted this identity, offered no independent rationale for the formula, proposing it only as an approximation to $\alpha_{(n_i)}$. Alexander [1] proposed (8) as a replacement for (2′), to be used when the M_i are presumed equal; his derivation implies equivalence of conditions. Formulas essentially the same as (7) and (8) were developed by Horst [18] for the analysis of ratings, though his development is strictly valid only for large n_i and n_p. He required that all raters 'measure the same function'—an extramathematical assumption like that of Spearman and Gulliksen. He required also that one set of ratings (wherein different raters rate different subjects) be equivalent to other such sets in means, variances, and between-set correlations.

---

* It is to be noted that this is the first place in our summary where sampling of persons is taken into account. All classical formulas were derived for populations rather than samples.

A rationale for (7) and (8) that successfully avoids equivalence assumptions was offered by Webster [39]. Webster assumes *only* that there is an infinite universe of tests (or raters, etc.) from which the tests to be applied to each individual is sampled at random, independently for each individual. That is, persons take different tests or are judged by different raters. Randomness of sampling guarantees that in the population the means, variances, and intercorrelations of scores will be equal. One set of such randomly generated data is equivalent to another even though the tests *individually* are not equivalent. Mean squares from a one-way analysis of variance lead to (7) and (8).

To summarize: The ratio of 'true' variance to observed-score variance is the intraclass correlation. There are two types of intraclass correlation, α and ρ^2_{MX}, corresponding respectively to the definition of the population observed-score variance as

$$\lim\,(n_p\rightarrow\infty)\;(1/n_p)\sum_p(X_{pi}-M_i)^2 \quad\text{and}\quad \lim\,(n_p,n_i\rightarrow\infty)\;(1/n_pn_i)\sum_p\sum_i(X_{pi}-M)^2$$

Previous derivations of α-type reliability formulas have assumed equal variances and intercorrelations among tests; in using the intraclass correlation for internal-consistency analysis, equivalence among part-or item-scores has been implicitly assumed. No equivalence assumption is required to use $\hat{\rho}^2_{MX}$ as a reliability estimate for data obtained by observing each individual under different randomly selected conditions.

Treatments for nonequivalent measures. Kelley [23] pointed out the possibility of defining reliability without requiring tests to be strictly equivalent. Both he and Cureton [10, 11], who developed this idea more explicitly, defined reliability as the squared correlation of true score M_p with score X_{pi} on a particular test i—we shall call this $\rho^2_{Mi}\cdot V_i$ and the regression slopes of X_{pi} on M_p were allowed to vary from test to test. For Kelley, the 'true score' for test i included all factors the test has 'in common with the general field' to be measured. Kelley offered no new formulas, but Cureton proposed ways of estimating a separate coefficient ρ^2_{Mi} for each of several tests representing the same field [10, pp. 11–12, 14, 17]. Cureton, assuming unit rank among the tests, in effect made a Spearman factor analysis of test scores and estimated the squared loading of each test on the general factor. His formulas never seem to have been adopted for the appraisal of reliability.

In the early 1930's Burt investigated, by means of factor analysis, the agreement among grades given to examination papers. Grades given by different judges obviously do not have uniform variances and intercorrelations. Burt's work was neglected both because it appeared only in an appendix to the main report on examinations [16] and because he turned shortly afterwards to analysis-of-variance models. His 1936 paper took the loading of the individual grader on the first factor among graders as an index of that grader's typicality, the first factor being defined by the centroid, simple-summation, or principal-component solution. This treatment is free of equivalence and unit-rank assumptions.

The formulation which will now be presented has a close affinity to many of the historical papers that have been reviewed; in particular, it includes features of Burt's 1936 work, of Kelley's, of Webster's, and of Spearman's 1904 paper. Prominent among the formulas to be recommended are intraclass correlations, but the nonequivalence of measures is taken specifically into account.

II. GENERALIZABILITY: BASIC CONCEPTS

An investigator asks about the precision or reliability of a measure because he wishes to generalize from the observation in hand to some class of observations to which it belongs. To ask about rater agreement is to ask how well we can generalize from one set of ratings to ratings by other raters. To ask about the reliability of an essay-examination grade is to ask how representative this is of grades that might be given to the same paper by other markers, or of grades on other papers by the same subject. Questions concerning generalizability are substantive, not purely methodological. They inquire about the postulated class of observations as well as about the measuring technique. 'Do ratings of likeability agree?' is a question in social psychology as much as it is a question about the appraisal of likeability. Spearman, asking whether he could generalize from a measure of touch discrimination to discrimination without regard to modality, was appraising the validity of 'discriminatory ability' as a theoretical construct for differential psychology.

Since a given measure may reasonably be generalized to many different universes, the investigator must specify the universe which is of interest to him before he can begin to study generalizability. This consideration is omitted from the classical theory, where it is implied that every test has *a* true score, belongs to only one family of parallel tests, and has only one 'reliability coefficient'. Guttman [15, p. 125f.] emphasizes this inadequacy in the theory:

> "Consider the following example of a series of 'parallel' tests. Let test 1 consist of but a single item: 'Write down all the words you can think of that begin with the letter *t*'. For a given population, and a given time limit, the score for each person is the number of words he writes down beginning with *t*.

There are at least two different directions in which one could go to construct tests parallel to this one. One direction is to vary *the letter involved*. For example, test 2 could be: 'Write down all the words you can think of that begin with *p* ', while test 3 could use instead the letter *d*, say. By adjusting the time limits, all three tests can be made to have the same mean. There seems no absolute barrier to their also having common variances and correlation coefficients. For our particular population, let us suppose the three tests are actually parallel, and that their common correlation coefficient is ·70. Then according to the .... theory, test 1 has reliability coefficient ·70.

Another direction in which we could have gone to construct tests parallel to test 1 is to vary the places of the letter, and not the letter itself. Thus, test 2 could be: ' Write down all the words you can think of in which the second letter is *t* ', and test 3 could ask for *t* as the *third* letter. Again, for our population, there is no physical bar to the tests turning out to be parallel. But this time, let us assume that the mutual intercorrelations turn out to be equal to ·60. Then test 1 has reliability ·60.

Therefore, test 1 has reliabilities ·70 and ·60 simultaneously, according to the theory of parallelism. It should be noted, too, that Gulliksen's nonalgebraic requirements are also satisfied simultaneously by each series. These additional requirements are ' the tests should contain items dealing with the same subject matter, items of the same format, etc.' "

We require the investigator to specify a universe of conditions of observation over which he wishes to generalize. *Conditions* is a general term referring to particular test forms or stimuli, observers, occasions or situations of observation, etc. (cf. [32] pp. 57 ff.). The observations within the universe may vary in more than one respect (e.g., for a measure of GSR in response to words selected from a long list, scores can be obtained for various words and electrode placements). We shall discuss only singly classified conditions, but the argument can be extended to multiple classifications.

The universe encompasses a finite or infinite number of conditions. We shall restrict this presentation to the infinite case, and shall also treat the conditions as unordered. No assumption is made about the content of the universe or about the statistical properties of the scores within conditions. The following requirements or assumptions are made:

1. The universe is described unambiguously, so that it is clear what conditions fall within the universe.

2. Conditions are experimentally independent; the person's score in condition i does not depend on the fact that he has, or has not been previously observed under other conditions.

3. Scores X_{pi} are numbers on an interval scale.

For each person p, the *universe score* M_p is defined as the mean of X_{pi} over all conditions in the universe.

To investigate the relation between the observed score and the universe score, a generalizability study (G study) is required. It will be advantageous to distinguish this from the decision (D) study. The D study provides data from which decisions about individuals or groups are made, while the G study is used to assess the measuring technique. This distinction is important because it is the decision data whose dependability is of interest; recognizing this aids the investigator in choosing among alternative formulas. Often, the same set of data is used for the analysis of generalizability as for decision making; in this case the G and D studies are the same. But the G and D studies may be separate, and may have different experimental designs. For example, G data on a method for appraising teacher behaviour may consist of simultaneous records made by two observers, but only one observer may collect the data for the main study relating teacher actions to pupil achievement.

In describing experimental designs we shall use the terms *matched* and *unmatched.* Data (for either the G or D study) are said to be *matched* when, for every i in the study, there is an X_{pi} for every person in the sample. When all persons take the same test, the resulting data are matched. Data are *unmatched* when the conditions of observation are selected independently for each person; save for chance duplications, each person is observed under different conditions. Ratings of university applicants by their high-school principals generally are unmatched with regard to rater; subject and rater are confounded in the experimental design. (We shall not deal with partial, incomplete matching). Where matched data are obtained under some condition i, we shall denote the score parameters (variance, covariance, correlation with universe score, etc.) by V_i, C_{Mi}, ρ_{Mi}, . . . respectively. For unmatched data, we represent the parameters by V_X, C_{MX}, ρ_{MX}, etc.

We now define the squared correlation $\rho^2_{M_pX_{pi}}$ as the coefficient of generalizability, which indicates how well one can generalize from the decision data to the universe score. When the decision data are matched on some condition i, this coefficient is ρ^2_{Mi}, and when they are unmatched, it is ρ^2_{MX}. In either case, the coefficient reports what proportion of variance in the set of X_{pi} (for the population) is linearly predictable from M_p. There may be a different ρ^2_{Mi} for each i in the universe. This definition of the coefficient is the same as that of Kelley and Cureton in all save our more explicit and perhaps less restricted definition of M_p.

One aim of a study of generalizability might be to estimate the specific coefficient ρ^2_{Mi} for a designated condition i. This is a reasonable objective where a certain condition will be used repeatedly in matched D studies; good estimates of ρ^2_{Mi}, $M_i - M$, and V_i for a certain observer, for example, would establish a personal equation for estimating M_p from the data he produces. In Spearman's study of discrimination, the specific coefficient for each test (e.g., colour discrimination) would indicate how well it represents the postulated general capacity to discriminate. More often, however, it is appropriate to inquire about the generalizability of whatever undesignated condition might be used in decision making. This is necessary when we do not know which condition will be used in collecting subsequent matched decision data, or when it has not been possible to include that condition in the G study. It is also appropriate when we want to judge the coherence of the universe as a whole, as in deciding whether likeability is a useful construct. We select $\underset{i}{E}\rho^2_{Mi}$ as the best single index of generalizability for matched data.

To our earlier assumptions regarding the nature of the universe and the scores within it, it is necessary to add an assumption regarding sampling. We assume that conditions used in the D study or the G study are randomly selected from the universe of conditions (except that, if a certain i is to be used in the D study, we may choose to include that as a fixed condition in the G study along with randomly sampled conditions). This random-sampling assumption is the distinctive feature of our argument. We also assume that in the G and D studies the persons observed are randomly sampled from the pertinent population. We let n_p represent the number of persons in the G study, n_i the number of conditions applied in the G study, and n_i' the number of conditions applied in the D study. It is assumed that n_i and n_i' are uniform for all persons, but n_i' need not equal n_i.

A G study may collect matched or unmatched data, the former providing more information. The matched G study with singly classified conditions provides estimates of V_M, V_X, EV_i, V_{M_i}, and the residual variance. This enables us to estimate $E\rho^2_{Mi}$, ρ^2_{MX}, and, in principle, the specific $\rho_{Mi}{}^2$ for any condition i used in the G study. The unmatched G study allows us to estimate V_M, V_X, and the within-persons variance; from these, ρ^2_{MX} can be estimated but not ρ^2_{Mi} or $E\rho^2_{Mi}$.

To make clear the extent to which we break away from restrictive assumptions, we may set up a model comparable to (4) but far more general. Scores are considered to be determined by the following factors.

> The first centroid factor of the covariances between conditions in the universe; the person's score on this factor is the universe score M_p.
>
> $F_1F_2, \ldots$; other centroid factors required to account for covariances between conditions. For convenience, scores X_{pF} on any factor are standardized with variance one.
>
> δ_i; residual variance after removal of the centroid factors. No restriction is placed on $V_{\delta pi|i}$, the variance of δ_{pi} with i fixed.

Now, introducing factor loadings $b_i\sigma_{M_p}$ and a_{Fi}, we may write

$$X_{pi} = M + b_i(M_p - M) + (M_i - M) + \sum_F (a_{Fi}X_{pF}) + \delta_{pi} \qquad (9)$$

The mean of a_{Fi} for any F is zero, and the mean of the b_i is one. It is now evident that we have discarded the unit-rank assumption and the Spearman assumption that the regression of X_{pi} on M_p is the same for all i. It follows that†

---

† John Hunter points out (personal communication) that for strict mathematical reasoning it is necessary to take these statements as *definitions* of variance and covariance rather than as inferences.

$$V_i = b_i^2 V_{M_p} + \sum_F a^2{}_{Fi} + V_{\delta_{pi|i}} \tag{10}$$

$$C_{Mi} = b_i V_{M_p} \text{ and } \rho^2{}_{Mi} = (b_i^2 V_{M_p} / V_i) \tag{11}$$

For unmatched data,

$$V_X = (V_b + 1) V_{M_p} + \underset{i}{E} \sum_F a^2{}_{Fi} + E_1 V_\delta + V_{M_i} \tag{12}$$

$$C_{MX} = V_{M_p} \text{ and } \rho^2{}_{MX} = V_{M_p} / V_X \tag{13}$$

III. ESTIMATE OF THE SPECIFIC $\rho^2{}_{Mi}$ FROM MATCHED G DATA

Suppose that condition i^* is to be used in subsequent D studies, and that it is included in the G study along with $n_i - 1$ other conditions randomly selected from the universe. Let i' be any condition other than i or i^*. Since $M_p = \underset{i}{E} X_{pi}$, $C_{Mi*} = \underset{i}{E} C_{i*i}$ and $V_{M_p} = \underset{i,i'}{E} C_{ii'}$ (where, for simplicity, we let E represent the average over all conditions or pairs of conditions in the universe). Then, from (11),

$$\rho^2{}_{Mi*} = C^2{}_{Mi*} / V_{M_p} V_{i*} = (EC_{i*i})^2 / EC_{ii'} V_{i*} \tag{14}$$

Since the conditions i (other than i^*) are randomly sampled, the G study permits unbiased estimates of the needed quantities provided the number of conditions in the universe is very large. Let $s_{ii'}$, represent a sample covariance. Then

$$\hat{E}C_{i*i} = \sum_i s_{i*i} / (n_i - 1) = \bar{s}_{i*i} \tag{15}$$

$$\hat{E}V_{M_p} = \sum_i \sum_{i'} s_{ii'} / (n_i - 1)\,(n_i - 2) = \bar{s}_{ii'} \tag{16}$$

$$\hat{V}_{i*} = s^2{}_{i*} \tag{17}$$

Substituting these estimates in (14),

$$\hat{\rho}^2{}_{Mi*} = (\bar{s}_{i*i})^2 / (\bar{s}_{ii'})(s^2{}_{i*}) \tag{18}$$

Letting T stand for the unweighted sum of n_i item scores, (18) becomes

$$\hat{\rho}^2{}_{Mi*} = (n_i - 2) s^2{}_{i*(T-i*)} / (n_i - 1)\,(s^2{}_{(T-i*)} - \sum_i s_i{}^2) s^2{}_{i*} \tag{19}$$

Formula (19) differs only slightly from formulas for a specific coefficient suggested by Burt in 1936. He correlated each condition with the first factor in the sample of conditions, while we estimate its correlation with the first factor in the universe. Furthermore, Burt standardized scores within conditions before making his analysis; that is, he referred the X_{pi} to the first factor in the correlation matrix rather than the first factor in the covariance matrix which is our M_p. Burt's formula (xxv) is otherwise comparable to our (19).

There are two drawbacks to the use of (19). Estimates are subject to extreme sampling fluctuations unless n_i is large or there is unit-rank among conditions. Second, the estimates tend to be slightly high, but this bias is small relative to the variable error. These conclusions are based on an unpublished numerical study by James Terwilliger.

Hence, although formula (19) has some theoretical interest, it is of little practical importance. The chief reason for desiring specific coefficients for various conditions is to identify which ones provide the best information about the universe score. This can be judged adequately from the easily computed $\rho^2{}_{Ti}$ (or $\rho^2{}_{(T-i)i}$). Though $\rho^2{}_{Mi^*}$ indicates more precisely than $\rho^2{}_{Ti}$ how well X_{pi^*} corresponds to M_p, it will rarely be practicable to use a sufficient number of conditions to estimate the magnitude of $\hat{\rho}^2{}_{Mi^*}$ very exactly.

Henrysson [17, formula 35] has suggested a formula closely similar to ours in another context. He considers that, for item analysis, one might wisely replace the usual item-test correlation by the estimated correlation of the item with the universe score. This would have particular value in comparing test items where, in preliminary work, different items are presented as parts of different tests. Using assumptions somewhat different from ours, Henrysson reaches a formula which differs from (18) only in that the average covariance in his denominator includes the s_{i^*i} as well as the $s_{ii'}$.

IV. VARIANCE ANALYSIS OF MATCHED G DATA

Discrepancy between $E\rho^2{}_{Mi}$ and V_{M_p}/EV_i. In previous formulations of reliability theory, it has been taken for granted—as a consequence of explicit or implicit assumptions—that for all i and i'

$$\rho^2{}_{Mi} = \rho_{ii'} = V_{M_p}/V_i \tag{20}$$

This is not necessarily true under our assumptions. It can be shown by means of the Cauchy-Schwartz inequality that V_{M_p}/EV_i is equal to or less than $E\rho^2{}_{Mi}$. We proceed on the assumption that the former, an intraclass correlation that can be estimated, is a satisfactory approximation to the latter. The discrepancy between the two is a complex function of the joint distribution of b_i and V_i. One noteworthy fact is that when all V_i are equal then $V_{M_p}/EV_i = E\rho^2{}_{Mi}$, regardless of whether correlations among conditions are equal or have unit rank. Preliminary calculations with hypothetical distributions show that the intraclass correlation is often a very good approximation to $E\rho_{Mi}{}^2$; a thorough computational study which has not yet been completed is expected to indicate under what conditions a good approximation is obtained.

Estimates of V_{M_p} and EV_i. Among 'models' for the analysis of variance components, only that of Cornfield and Tukey appears suitable for our problem. We wish to consider both p and i as randomly sampled and to allow for correlated interactions. Cornfield and Tukey [7] describe a 'pigeonhole' model which dates back to an unpublished (1949) memorandum by Tukey. Assuming random sampling of p and i and allowing correlated interactions, they derive expressions for expected mean squares which are essentially the same as those more commonly reached by assuming interaction components to be independent random variables.

The model, say Cornfield and Tukey, has " sufficient generality and flexibility that the necessary assumptions concern only the selection of the levels of the factors (the behavior of the experimenter) and not the behavior of what is being experimented upon ". In a two-way classification problem, it is assumed that there exist pigeonholes arranged in rows and columns, with numerous elements (potential observations) in each pigeonhole. Subsets of n_p rows and n_i columns are selected, randomly and independently but without replacement; and in each of the $n_i n_p$ pigeonholes where these rows and columns intersect a sample of n values of X_{pi} is drawn. The additive model is written in the form

$$X_{pik} = \begin{Bmatrix}\text{general}\\ \text{mean}\end{Bmatrix} + \begin{Bmatrix}\text{row}\\ \text{effect}\end{Bmatrix} + \begin{Bmatrix}\text{column}\\ \text{effect}\end{Bmatrix} + \begin{Bmatrix}\text{cell}\\ \text{effect}\end{Bmatrix} + \text{residual} \tag{21}$$

The cell effects are ' tied ' to the corresponding row and column effects in the sense that when a particular person and condition have been selected all three values are determined [7, p. 918]. The cell effect represents the pigeonhole mean, after subtracting the row and column effects and the general mean. The residual represents the within-pigeonhole deviations, which are to be regarded as independent random variables; there may be a different residual variance within each pigeonhole. Our eqn. (9) corresponds to this model, as is evident if we divide δ_{pi} into two portions so that we can rewrite (9) as

$$X_{pik} = M + \underset{\substack{\text{row}\\ \text{effect}}}{(M_p - M)} + \underset{\substack{\text{column}\\ \text{effect}}}{(M_i - M)} + \underbrace{(b_i - 1)(M_p - M) + a_{Fi}X_{pF} + \delta'_{pi}}_{\text{cell effect}} + \underset{\text{residual}}{\delta''_{pik}} \tag{9'}$$

Cornfield and Tukey represent the row, column, cell, and within-cell components of variance by σ_R^2, σ_C^2, σ_{RC}^2, and σ_E^2, respectively. If the population of persons and the universe of conditions are infinite, the expected mean squares from a two-way analysis of variance with n entries per cell are as follows:

$$\left.\begin{aligned} &\text{EMS rows} &&= \sigma_E^2 + n\sigma_{RC}^2 + nn_i\sigma_i\,\sigma_R^2\\ &\text{EMS columns} &&= \sigma_E^2 + n\sigma_{RC}^2 + nn_p\sigma_C^2\\ &\text{EMS interaction} &&= \sigma_E^2 + n\sigma_{RC}^2\\ &\text{EMS error} &&= \sigma_E^2 \end{aligned}\right\} \tag{22}$$

In test data with one observation per cell, it is impossible to separate the interaction component of variance from the within-cell residual. With $n=1$, the eqns. become

$$\left.\begin{aligned} &\text{EMS rows (persons)} &&= \sigma_E^2 + \sigma_{RC}^2 + n_i\sigma_R^2\\ &\text{EMS columns (conditions)} &&= \sigma_E^2 + \sigma_{RC}^2 + n_p\sigma_C^2\\ &\text{EMS residual} &&= \sigma_E^2 + \sigma_{RC}^2 \end{aligned}\right\} \tag{23}$$

The expected variance within any column is

$$EV_i = \sigma_R^2 + \sigma_{RC}^2 + \sigma_E^2 \tag{24}$$

We may replace $\sigma_{RC}^2 + \sigma_E^2$ with σ_e^2—where, as in (4), e_{pi} is the cell effect plus the residual, but is not now regarded as the same random variable in every cell. We rewrite (23), recognizing that σ_R^2 is V_{M_p} and σ_C^2 is V_{M_i}:

$$\left.\begin{array}{ll}\text{EMS}_p\text{ (persons)} & = n_i V_{Mp} + \sigma_e^2 \\ \text{EMS}_i\text{ (conditions)} & = n_p V_{Mi} + \sigma_e^2 \\ \text{EMS}_r\text{ (residual)} & = \sigma_e^2\end{array}\right\} \qquad (25)$$

Consequently, from the matched G study (cf. (5))

$$\left.\begin{array}{l}\hat{V}_{Mp} = (1/n_i)(MS_p - MS_r) \\ \hat{V}_{Mi} = (1/n_p)(MS_i - MS_r) \\ \hat{V}_e = MS_r\end{array}\right\} \qquad (26)$$

and

$$E\hat{V}_i = \hat{V}_{Mp} + \hat{V}_e = (1/n_i)(MS_p + (n_i - 1)MS_r) \qquad (27)$$

Suppose that the score from which a decision will be made is the average of scores on n_i' randomly sampled conditions. Then, from the usual equation for the variance of a mean, the expected variance of the decision data is

$$EV_{i_{(n_i')}} = V_{M_p} + (1/n_i')V_e \qquad (28)$$

The estimates from (26) may be substituted. All the foregoing formulas can readily be adapted for a universe with a finite number of conditions.

The intraclass correlation. To obtain the intraclass correlation we substitute unbiased estimates for the variances,† so that we have

$$\hat{\alpha}_{(n_i')} = \hat{V}_{M_p}/E\hat{V}_{i(n')} = n_i'(MS_p - MS_r)/(n_i'MS_p + (n_i - n_i')MS_r) \qquad (29)$$

This eqn. holds whether the composite score is obtained by averaging or by adding the n'_i part scores. If $n'_i = n_i$, whether or not the same data are used for the G and D studies,

$$\alpha_{(n_i)} = (MS_p - MS_r)/MS_p \qquad (30)$$

which is the Hoyt formula. In internal-consistency studies one may obtain $\hat{\alpha}_{(n_i)}$ by KR20 or the computational forms (1) and (2) given earlier. This estimation formula for the intraclass correlation has now been justified without assuming equivalence, unit-rank, or uniformity of the within-cell variances σ_e^2. Without equivalence, however, we can only say that the intraclass correlation α is approximately equal to $E\rho_{ii'}$ and a more-or-less close lower bound to $E\rho^2_{Mi}$.

---

† The studies of Olkin and Pratt [29] indicate that the estimate is less—by a negligible amount—than the population value of the ratio. They, however, consider only errors arising from sampling of persons; sampling of conditions adds some further bias. Work by Walter Kristof, not yet published, suggests a small correction for errors arising from person sampling which converts $\hat{\alpha}$ into an unbiased estimate of α. He, however, restricts himself to the case where conditions are strictly equivalent except in means, and scores have a multinormal distribution.

The general formula (29) yields the following equation for any two values of n_i (say n_1 and n_2, where $n_2/n_1 = k$):

$$\hat{\alpha}_{(n_2)} = k\hat{\alpha}_{(n_1)}/(1+(k-1)\hat{\alpha}_{(n_1)}) \tag{31}$$

In form, this is identical to the Spearman-Brown formula; using the intraclass rather than the interclass correlation makes it possible to derive the formula from a much less restrictive model.

Formula (29) or the special case (30) can be recommended whenever matched data are employed in the G and D studies. Consider several specimen problems.

(*a*) Six judges rate certain protocols in a G study. In the D study, some judge presently not designated but presumed to be randomly selected from the same universe will rate all protocols. Then (29) should be used with $n' = 1$.

(*b*) In the situation described in (*a*), one of the six judges of the G study will be used in the D study. Again (29) with $n' = 1$ is recommended. While it would be possible to estimate ρ^2_{Mi} for the particular judge, by means of (19), the variable error makes the result untrustworthy.

(*c*) A test of 25 items is administered in the G study. The same test, or some other set of 25 items from the same universe, will be used in the D study. Then (30) should be used, with $n_i = 25$. (Each item, rather than each test, is regarded as a ' condition '.)

(*d*) Two judges are used in the G study. How many judges will be used in the D study remains to be decided. One should apply (29), with n'_i unknown. The value of $\hat{\alpha}$ for various n' can be calculated, or one can solve for the n' required to produce any desired $\hat{\alpha}$.

The critical assumption behind these applications is, as elsewhere, that conditions (judges, items, etc.) are randomly sampled from the same universe in both G and D studies.

While it is customary to seek a coefficient in reliability studies, greater importance attaches to the estimates of variance components, which are useful in understanding the origins of error in planning efficient experimental procedures. The investigator should report his estimates of the separate variance components V_e, V_{M_p}, and V_{M_i}; he can then calculate $\hat{E}V_i$ and $\hat{\alpha}$ for one or more values of n_i' or can leave this to the reader. The estimated variances convey more information than the coefficient alone. The magnitude of V_{M_i}, for instance, sheds light on the importance in the D study of observing all persons under the same condition; the greater this variance, the more important such a control becomes. Reporting all variances also helps in examining the error of measurement, the error of a group mean, the error of an item mean, etc.

Reporting variance components is increasingly advantageous as conditions are classified with respect to more facets. A multifacet G study sheds light on the efficiency of many alternative experimental designs with different crossing or nesting of facets, and different numbers of conditions. The traditional emphasis on a single coefficient for a given measuring device and sample of persons is appropriate only where the design of the G study is to be used in all D studies. But the separate variance estimates from the G study shed light on how the technique may most profitably be applied in the D study. Some discussion of such use of G studies is to be found in Lindquist [25, pp. 357 ff.]; this point of view should receive greater emphasis than it has in the training of investigators and test users.

Estimates for D studies in which data are unmatched. Components of variance in unmatched D data may be estimated from the matched G study, as was pointed out by Rajaratnam [30]. As the sample of persons is increased, the score variance in such D data will always approach V_X, since conditions are sampled independently for each person.

$$V_{X(1)} = \underset{i,p}{E}\,(X_{pi} - M)^2 = \underset{i,p}{E}\,(X_{pi} - M_i)^2 + E(M_i - M)^2 = EV_i + V_{M_i} \tag{32}$$

Substituting from (26) and (27),

$$\hat{V}_{X(1)} = (1/n_i)[MS_p + (n_i - 1)MS_r] + (1/n_p)[MS_i + MS_r] \tag{33}$$

The estimate of V_{M_p} from (26) is divided by (33) to obtain

$$\hat{\rho}^2_{MX(1)} = \hat{V}_{M_p}/\hat{V}_X = [MS_p - MS_r]/[MS_p + (n_i/n_p)MS_i + (n_in_p - n_i - n_p/n_p)MS_r] \tag{34}$$

More generally, where decisions are based on a composite of n_i' observations,

$$\hat{\rho}^2_{MX(n_i')} = n_i'[MS_p - MS_r]/[n_i'MS_p + (n_i/n_p)MS_i + (n_in_p - n_i'n_i - n_p/n_p)MS_r] \tag{35}$$

Again, the Spearman-Brown relationship holds:

$$\hat{\rho}^2_{MX(n_2)} = k\hat{\rho}^2_{MX(n_1)}/(1 + (k-1)\hat{\rho}^2_{MX(n_1)}) \tag{36}$$

This set of formulas is pertinent, for example, when, for purposes of evaluating a technique, two judges rate all of a set of protocols by means of a scoring guide, but in subsequent use of the technique it is anticipated that a protocol may be scored by any of a large group of scorers. One might similarly control observers, occasions, or some other facet in a pilot study, and not exercise the same control in routine use of a method. Lord [27] suggests use of (34) where each person takes one of a limited set of test forms, and the investigator proposes to disregard what form the person took when interpreting his score.

V. VARIANCE ANALYSIS OF UNMATCHED G DATA

Where each person is observed under different conditions in the G study, only a one-way analysis is possible. We recapitulate here the argument of Webster [39], assuming the number of observations n_i to be the same for all persons. The column, cell, and within-cell components of (22) being inseparable we write V_ϵ for the within person EMS, ϵ_{pi} being the 'error of measurement' $X_{pi} - M_p$. In place of (23) and (25) we have

$$\text{EMS rows (persons)} = \sigma_E^2 + \sigma_{RC}^2 + \sigma_C^2 + n_i\sigma_R^2 = n_i V_{M_p} + V_\epsilon$$

$$\text{EMS within persons} = \sigma_E^2 + \sigma_{RC}^2 + \sigma_C^2 = V_\epsilon \tag{37}$$

Consequently,

$$\hat{V}_M = (1/n_i)(MS_p - MS_{wp}) \tag{38}$$

$$\hat{V}_{\epsilon_{(1)}} = MS_{wp} \tag{39}$$

Considering decision data based on averages of n'_i items,

$$\hat{V}_{\epsilon_{(n_i')}} = (1/n_i')MS_{wp}$$

$$\hat{V}_{X_{(n_i')}} = (1/n_i)MS_p + (1/n_i' - 1/n_i)MS_{wp} \tag{40}$$

$$\hat{\rho}^2_{MX_{(n_i')}} = n_i'(MS_p - MS_{wp})/(n_i' MS_p + (n_i' - n_i)MS_{wp}) \tag{41}\dagger$$

For $n_i' = n_i$,

$$\hat{\rho}^2_{MX_{(n_i)}} = (MS_p - MS_{wp})/MS_p \tag{42}$$

The Spearman-Brown formula (36) applies to $\hat{\rho}^2_{MX}$. (42) is identical to (8) and if $n_i' = 1$, to (7). Since $V_X \geq EV_i$, $\alpha \geq \rho^2_{MX}$.

A quite distinct line of argument leads to the application of (41) and its special case KR21 to *matched* G data, as a lower bound to the logically appropriate formula. Since in samples of reasonable size $MS_{wp} \geq MS_r$, if (41) is applied to matched data it will give a lower bound to the $\hat{\alpha}_{(n')}$ that would be obtained by applying (29) to those data. For n_p large, (41) also gives a lower bound to the $\hat{\rho}^2_{MX(n')}$ that would be obtained by applying (35). (This is a modification of the argument by which Kuder and Richardson justified KR21).

Formula (41) and its variants are to be recommended in such studies as the following:

(*a*) Children are observed on the playground, the time when each child is observed being randomly determined. In the G study, each child is observed three times; in the D study, once. Formula (41) is used.

(*b*) Three former employers of each applicant are asked to fill out a rating form. The same data are used for the G and D study, decisions being made on the basis of the average rating. It is rare that any rater rates more than one applicant. Formula (42) is used.

VI. Discussion

Assumptions used. The importance of this reformulation obviously does not lie in any new formulas or procedures for the study of tests. The formulas recommended are identical to or slightly generalized versions of those already

---

† Again, we note that a ratio of estimates is a biased estimate of the ratio; but since the number of *different* conditions used is nearly equal to n_p, the Olkin-Pratt computations give assurance that this bias is negligible, compared to the sampling variance of the estimate.

in use. We trust, however, that obscurities and inconsistencies in the choice of formulas are eliminated by this development.

By applying the mathematical rationale of Cornfield and Tukey we show how the intraclass correlation may be used in 'reliability' analysis with no assumption save that persons and conditions are randomly and independently sampled from a population and universe. The chief formulas are summarized in Table 1; for estimates for unmatched D data from matched G data, see (32)–(36) above. We find that for matched data the intraclass correlation $\alpha = V_{M_p}/EV_i$ is a lower bound to the coefficient of generalizability as we define it. For unmatched decision data, the intraclass correlation $\rho^2_{MX} = V_{M_p}/V_X$ equals the desired coefficient of generalizability.

The intraclass correlation—whether for matched or unmatched data—changes according to the Spearman-Brown formula as the number of observations changes. This frees that formula of its dependence on strong equivalence assumptions.

The conceptual separation of the generalizability and decision studies clarifies the choice of formulas. Previous writers have had difficulty in stating clearly when $\hat{V}_{M_i}$ should be taken into account in the intraclass correlation. We find that the choice of formula depends on the design of the G study and the anticipated design of the D study.

TABLE 1. CHIEF FORMULAS

Data in G Study (n_i observations)	Matched	Unmatched
Data in D Study (n_i' observations averaged)	Matched	Unmatched
Variance estimates for D study:		
Universe score	$\frac{1}{n_i}(MS_p - MS_r)$	†
Error	$\frac{1}{n_i'}MS_r$	†
Observed variance	$\frac{1}{n_i}MS_p + \left(\frac{1}{n_i'} - \frac{1}{n_i}\right)MS_r$	†
Intraclass correlation:		
n_i' variable	$\frac{n_i'(MS_p - MS_r)}{[n_i'\,MS_p + (n_i - n_i')MS_r]}$	†
$n_i' = n_i$	$\frac{MS_p - MS_r}{MS_r}$ ‡	† §
$n_i' = 1$	$\frac{MS_p - MS_r}{[MS_p + (n_i - 1)MS_r]}$	†

†Same as in first column except that MS_{wp} replaces MS_r.
‡KR20 or our (2) may be used as a computing form.
§KR21 may be used as a large-sample computing form.

Definition of coefficient. There is no doubt that the coefficient ρ^2_{Mi} or ρ^2_{MX} serves the traditional purposes of a reliability coefficient. It indicates the proportion of observed variance linearly predictable from universe score and can be used in computing a confidence interval for true score or a regression estimate of it, and in correcting correlations for attenuation. It is rarely practicable to estimate the separate ρ^2_{Mi}. Our decision to represent these many specific coefficients by their mean is a somewhat arbitrary choice. $E\rho^2_{Mi}$, the average proportion of variance arising from universe score, would not necessarily be as satisfactory as $E\rho_{Mi}$ in regression equations or corrections for attenuation. ($E\rho_{Mi}$ is less than $(E\rho^2_{Mi})^{1/2}$.) There is, however, no evident alternative to $(\hat{\alpha})^{1/2}$ as an estimator of $E\rho_{Mi}$.

A somewhat different question may be raised about the choice of M_p as universe score. An alternative would have been to use the average of standardized scores, m_p, in which the observations are weighted inversely by the standard deviations. M_p and m_p will have a very high correlation, since one is the first centroid of the covariance matrix and the other that of the correlation matrix. They tend to differ, however, as the correlation of the σ_i with the ρ_{mi} departs from zero. Conditions having relatively large standard deviations have greatest weight in determining M_p. The specific ρ^2_{Mi} therefore will not necessarily equal the ρ^2_{mi}, and the expectancies may differ slightly. If the universe score is to be thought of in absolute terms (for example, in measuring vocabulary size or breadth of acquaintance), M_p appears to be the better index; but if the universe score is to represent the mean of *relative* standing (in, for example, some quality rated by many judges), m_p would be preferred. Given matched G data for reasonably large n_p, one can divide each score by the s.d. for that condition and then apply the formulas of Table 1 to estimate V_{m_p} and $E\rho^2_{mi}$ or $E\rho^2_{mX}$. In such a calculation, $\hat{\alpha}_{(1)}$ is an estimate of $E\rho_{ii'}$. (The error resulting from use of sample s.d.'s in standardizing will be negligible). With an unmatched G study there is no alternative to using ρ^2_{MX} as an estimate of ρ^2_{mX}.

Interdependence of coefficient and universe. The reinterpretation of 'reliability' theory as a theory of generalizability removes many confusions from the application of measurement theory. The semantic problems of interpreting 'reliability', 'true score', and 'error' reduce to mere matters of syntax when we introduce the word 'generalizability'. To speak of the generalizability of a measure is obviously an incomplete statement until the speaker indicates what construct is being generalized to; he is forced to be explicit about what has often been implicit and therefore lost from sight. The so-called error of measurement becomes a discrepancy between the measurement and a universe score, and the question, 'What universe?' follows naturally. The person must declare what universe he has in mind: e.g., 'the 26 fluency tasks calling for construction of six-letter words with any specified initial letter'. Since there are many universes to which the test might be referred, the one reported coefficient does not pretend to answer all pertinent questions about the representativeness of the score.

It will be worthwhile to give examples of the confusions we hope to avoid by our terminology. An investigator claims that each of several laboratory observations represents a distinct temperamental variable. To support this claim he compares their intercorrelations, which are in the neighbourhood of 0·40, with their 'reliability coefficients' of 0·80 or so. There apparently is ample variance to support the specific-factor interpretation. On closer examination, we find that his 'reliabilities' are correlations between observers who score the same performance. But the claim that score on Task A represents temperament factor I is a generalization seemingly covering all forms of Task A and all occasions within a considerable time span, as well as a generalization over observers. Hence the correlation of A with another task should be compared with an estimate of $\rho^2_{M_A A}$ where M_A is the mean over a universe in which stimuli and occasions vary as well as observers. This coefficient would surely be much lower than 0·80, and might leave no room for a specific factor.

A slightly different example comes from a study of psychotherapy. Before and after treatment, a professional worker observed and interviewed the patient and described in a protocol his status at that time. To appraise adjustment quantitatively, the investigator asked judges to rate the protocol on a seven-point scale. He wishes to decide whether the patient made significant changes between interviews, and suggests that error of measurement be estimated from the agreement of two judges rating the same protocol. But this relates to the discrepancy between one rating and the mean for a universe of ratings of the same protocol. Here again the 'reliability coefficient' greatly overestimates the adequacy of the measures because too narrow a universe is represented in the replications. The investigator needs an estimate covering the universe of protocols that might be prepared by different *observers* who see the patient at different *moments* during the pretreatment period (and similarly for a post-treatment universe).

Our rationale requires the investigator to start his G study by defining the universe that interests him, and then to observe under two or more independently selected conditions within that universe. The calculations tell him how well the observed scores represent the universe scores. Since the the universe is a construct that he introduces because he thinks it has explanatory or predictive power, an investigation of generalizability is seen to be an investigation of the 'construct validity' of the measure [9]. Thus the theory of 'reliability' and the theory of 'validity' coalesce; the analysis of generalizability indicates how validly one can interpret a measure as representative of a certain set of possible measures. This point of view was made explicit by Tryon [37], who described the reliability coefficient as a measure of the 'domain validity' of a test. Though we supplement Tryon's argument at several points, the only major difference between us is that Tryon restricted the universe to a set of measures having equal variances and equal intercorrelations.

What universe is to be selected for attention depends on the hypotheses of the investigator or his practical purposes. To regard an applicant's score on a typing test as favourable is to forecast success in any task belonging to some particular universe. The investigator might want to generalize from a specimen of typing on typical business correspondence over other specimens of similar correspondence, if the subject is being considered for business employment. He might also wish to generalize over typewriting machines, if the offices to be staffed do not have uniform equipment. The question of generalizability from business copy to other copy arises if some vacancies call for scientific or literary typing. As a second example, consider playground observations of aggression. Time samples of behavior are taken when children are engaged in free play. An investigator with one theory might regard the score as representing the child's general aggressiveness in all interpersonal relations. Another, regarding home and school as quite different behavioral environments, might expect to generalize only over school situations. A third might expect substantial generalizability only if he narrows the universe to free-play situations within the school where the teacher maintains a certain defined supervisory attitude. These differences, it is obvious, pose empirical questions; a *G* study might well be designed to determine how confidently one can generalize over any of the universes mentioned. But there remains room for individual preferences in using the facts. An investigator may prefer the simplicity of a single score for aggressiveness to having a multiplicity of scores representing subuniverses, even if generalizability is much greater within subuniverses.

The possibility of defining the universe in different ways is well illustrated in the Campbell-Fiske proposal [5] for multimethod studies. They propose that a measure of (say) introversion should be accepted only if it agrees with other measures employing the same technique (e.g., self-rating) and also with measures by other techniques (e.g., observation). This is clearly an appropriate requirement if the measure is to represent the universe of all possible introversion measures, i.e., is intended as a measure of introversion without regard to technique. But if the investigator confines his interpretation to 'self-reported introversion' it is necessary to show only that the score estimates the expectancy over questionnaires and self-ratings; one might even have theoretical grounds for expecting that the score would *not* reflect introversion as judged from a projective technique or field observation.

The correction for attenuation provides a particularly good illustration of the value of designating clearly the universe under consideration. Suppose that we are predicting ability in English composition. The criterion may be the mark assigned to a specimen composition on an assigned topic, at the end of some training programme. The correlation of the predictor with that mark would be the validity coefficient, and the usual formula would 'correct' this for 'error in the criterion'. But this correction should estimate the correlation of the predictor with the universe score on the criterion. What universe? There are many possibilities: the universe of marks assigned by scorers to this composition, the universe of compositions that might be written on this topic in many trials, the universe of compositions written on different topics of the same general sort, etc. For each of these universes there is a different coefficient of generalizability and a different corrected validity. Each of these validities has its own meaning, hence it is highly important to identify the universe of greatest interest.

Designating a coefficient as 'an estimate of the expected coefficient of generalizability over (e.g.) test forms and trials' permits us to dispense with the unwieldy nomenclature of the *Technical Recommendations for Psychological Tests* [36]—viz., 'coefficient of stability', 'coefficient of equivalence', and 'coefficient of internal consistency'. Any idea that could be expressed by naming coefficients distinctively can be expressed more precisely by designating the universe to which each coefficient refers.

The coefficient as an indication of universe homogeneity. If a set of items is a sample from a homogeneous universe, one can use the observed scores with great confidence that any other sample from the same universe would give the same information. Among the studies that have explored methods of assessing homogeneity we need only mention Walker's pioneer work on the 'unique answer pattern' [38], Guttman's scale theory [35], and Loevinger's monograph [26].

The coefficient of generalizability (with $n_i'=1$) is for most purposes an adequate indicator of the homogeneity of the universe. A small coefficient implies that one could hope for appreciably closer relation between observation and interpretation if he narrowed or redefined the universe of generalization. Thus the coefficient helps to estimate the potential value of the proposed universe as a construct. Second, the coefficient tells how large a sample is required to estimate the universe score with any desired accuracy. Assume that there is no practical barrier to increasing or decreasing n_i' for the decision data. Then, no matter how heterogeneous or arbitrary the universe, so long as $V_M>0$, it is possible to adjust n_i' to bring the coefficient of generalizability to any desired magnitude. The greater the coefficient for $n_i'=1$, the smaller the sample required.

For many purposes, a heterogeneous universe may be quite satisfactory. 'Likeability', for example, can predict success in practical tasks such as teaching and military leadership. The fact that raters are somewhat heterogeneous in their judgments is primarily significant in indicating how many raters are needed to obtain a stable composite. To put it another way: single raters are not homogeneous, but averages over six raters may be quite satisfactorily homogeneous. While in principle more perfect prediction could be obtained by dividing raters into subsets with homogeneous preferences, there are likely to be good reasons for not introducing this complication unless we are developing a theory to explain preferences.

Objections to the random-sampling model. The interpretation of a psychological observation as a sample from a universe is as old as Galton. There have been objections, however, to using this as a model for test theory. The objector points out (i) that the universes to which one might refer are usually vaguely defined and not denumerable (e.g., what is the universe of all possible questions about symptoms of introversion?) and (ii) that strict random sampling from a pool of items or judges rarely occurs. We do not regard these difficulties as grave. With regard to the former, it is chiefly important that the universe be so well defined that different readers can agree in classifying particular instances as falling within or not falling within the universe. Thus, if it is claimed that scores obtained by a certain procedure are generalizable over all qualified scorers, the specification must be elaborate enough that we can decide whether graduates of a certain training programme are ‘ qualified ’. This is no more than a requirement for reporting experimental conclusions unambiguously; whether some universe is designated too vaguely can be checked by testing readers' agreement in applying the specification. The absence of true random sampling from a pool of items or judges is unfortunate but no more so than in the ubiquitous studies that make statistical inferences from persons who are not chosen in a strictly random fashion.

Sometimes a stratified-sampling model describes the process of test construction better than a random model. When several composites are made by sampling items according to some stratification plan (with random sampling within each stratum), the composites themselves are random samples from the universe of such composites. Therefore, $\hat{\alpha}_i$ is a suitable formula if more than one stratified composite is applied in the G study. Internal-consistency analysis of scores should take the sampling plan into account. Stratified versions of the formulas of this paper can be developed by a slight modification of the rationale given here; the stratified formulas comparable to (30) and (44) have been given by Tryon [37, p. 238] among others.

VII. SUMMARY

This paper has offered a theoretical rationale within which numerous lines of thought about reliability can be combined and reinterpreted. The investigator is presumably interested in generalizing from one observation to other observations that are members of the same universe. The universe consists of observations that might be made under various conditions; in the present paper, attention is confined to universes classified with respect to a single facet (e.g., test forms, observers, or occasions). The universe must be unambiguously defined, but it is not necessary that the universe be homogeneous in any other sense. A study of generalizability (G study) is conducted by observing persons under n_i conditions assumed to be randomly representative of the universe. A D study with n_i' conditions provides the data from which decisions will be made about individuals. Data may be matched (with all persons observed under the same conditions) or unmatched (conditions for each person being sampled independently).

With an unmatched G study, the one-way analysis of variance provides information about unmatched decision data, yielding estimates of the variance of universe scores, the expected variance of the difference between universe scores and observed scores, and the expected observed variance. The ratio of universe-score variance to observed variance is the coefficient of generalizability ρ^2_{MX} for unmatched D data. This intraclass correlation closely resembles the Horst generalized reliability coefficient and the KR21 formula. From matched G data it is theoretically possible to estimate a specific coefficient ρ^2_{Mi} for any condition included in the G study; the estimate is subject to a large sampling error and is not recommended unless n_i is quite large. A two-way analysis of variance leads to estimates of the universe-score variance and of the expected error and observed variances for both matched D data and unmatched D data. For matched D data, we estimate the intraclass correlation V_{Mp}/EV_i, here designated α; one special case of the estimation formula is KR20. α is a lower bound to the coefficient of generalizability $E\rho^2_{Mi}$. The intraclass correlation changes with n_i' according to the Spearman-Brown formula.

The Cornfield-Tukey rationale is used as a basis for deriving variance estimates. These estimates are unbiased even when conditions are not equivalent in the classical sense; we have specifically allowed for unequal variances, for unequal intercorrelations, and for intercorrelations with rank greater than one. The intraclass correlations are biased to some slight extent. Only at the step where the variance ratio α is taken as an approximation—as well as a lower bound—to $E\rho^2_{Mi}$ is there a serious gap in the rationale.

The most important formulas are collected in Table I which shows how the choice of formula depends on the design of the G and D studies. It is recommended that the person conducting a study of generalizability report all variance estimates as well as the coefficient appropriate to his D study; this becomes especially important in multi-facet G studies. Any coefficient should be accompanied by a careful description of the universe of generalization and the design of the D study to which it applies.

Our formulation, taken in combination with the discussions of Lindquist [25], Burt [4], and Tryon [37], removes much of the confusion from so-called reliability theory. It shows the consequence of applying formulas previously derived under equivalence assumptions to non-equivalent conditions, clarifies where various alternative formulas are to be used, and makes clear that different experimental procedures for studying 'reliability' relate the observed score to different universes of generalization.

REFERENCES

[1] ALEXANDER, H. W. (1947). The estimation of reliability when several trials are available. *Psychometrika*, XII, 79–99.

[2] BROWN, W. (1910). Some experimental results in the correlation of mental abilities. *Brit. J. Psychol.*, III, 296–322.

[3] BURT, C. (1945). The reliability of teachers' assessments of their pupils. *Brit. J. educ. Psychol.*, XV, 80–92.

[4] BURT, C. (1955). Test reliability estimated by analysis of variance. *Brit. J. statist. Psychol.*, VIII, 103–118.

[5] Campbell, D. T. and Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol. Bull.*, LVI, 81–105.

[6] Cast, B. M. D. (1939). The efficiency of different methods of marking English compositions. Part I. *Brit. J. educ. Psychol.*, IX, 257–269.

[7] Cornfield, J. and Tukey, J. W. (1956). Average values of mean squares in factorials. *Ann. math. Statist.*, XXVII, 907–949.

[8] Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, XVI, 297–334.

[9] Cronbach, L. J. and Meehl., P. E. (1955). Construct validity in psychological tests. *Psychol. Bull.*, LII, 281–302.

[10] Cureton, E. E. (1931). Errors of measurement and correlation. *Arch. Psychol.*, XIX, No. 125, 1–63.

[11] Cureton, E. E. (1958). The definition and estimation of test reliability. *Educ. and Psychol. Meas.*, XVIII, 715–738.

[12] Ebel, R. L. (1951). Estimation of the reliability of ratings. *Psychometrika*, XVI, 407–424.

[13] Gulliksen, H. (1950). *Theory of Mental Tests.* New York: John Wiley and Sons Inc.

[14] Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika*, X, 255–282.

[15] Guttman, L. (1953). A special review of Harold Gulliksen, *Theory of Mental Tests*. *Psychometrika*, XVIII, 123–130.

[16] Hartog, P. J. and Rhodes, E. C. (1936). *The Marks of Examiners.* London: Macmillan.

[17] Henrysson, S. (1963). Correction of item-total correlations in item analysis. *Psychometrika*, XXVIII, 211–218.

[18] Horst, P. (1949). A generalized expression for the reliability of measures. *Psychometrika*, XIV, 21–31.

[19] Hoyt, C. (1941). Test reliability estimated by analysis of variance. *Psychometrika*, VI, 153–160.

[20] Jackson, R. W. B. (1939). Reliability of mental tests. *Brit. J. Psychol.*, XXIX, 267–287.

[21] Jackson, R. W. B. and Ferguson, G. A. (1941). *Studies on the reliability of tests.* Bull. 12, Dept. of Ed. Res. Toronto: U. of Toronto.

[22] Kaitz, H. B. (1945). A note on reliability. *Psychometrika*, X, 127–131.

[23] Kelley, T. L. (1924). Note on the reliability of a test: A reply to Dr. Crum's criticism. *J. educ. Psychol.*, XV, 193–204.

[24] Kuder, G. F. and Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika*, II, 151–160.

[25] Lindquist, E. F. (1953). *Design and Analysis of Experiments in Education and Psychology*. Boston: Houghton-Mifflin Co.

[26] Loevinger, J. (1947). A systematic approach to the construction and evaluation of tests of ability. *Psychol. Monogr.*, LXI, No. 4.

[27] Lord, F. M. (1959). An approach to mental test theory. *Psychometrika*, XXIV, 283–302.

[28] Medley, D. M. and Mitzel, H. E. (1963). Measuring classroom behavior by systematic observation. Chapter 6 of *Handbook of Research on Teaching*, N. L. Gage (ed.), pp. 247–328. Chicago: Rand-McNally.

[29] Olkin, I. and Pratt, J. W. (1958). Unbiased estimation of certain correlation coefficients. *Ann. math. Statist.*, XXIX, 201–211.

[30] Rajaratnam, Nageswari (1960). Reliability formulas for independent decision data when reliability data are matched. *Psychometrika*, XXV, 261–271.

[31] Rulon, P. J. (1939). A simplified procedure for determining the reliability of a test by split halves. *Harvard Educ. Rev.*, IX, 99–103.
[32] Scott, W. and Wertheimer, M. (1962). *Introduction to Psychological Research.* N.Y.: John Wiley and Sons, Inc.
[33] Spearman, C. (1904). The proof and measurement of association between two things. *Amer. J. Psychol.*, XV, 72–101.
[34] Spearman, C. (1910). Correlation calculated from faulty data. *Brit. J. Psychol.*, III, 271–295.
[35] Stouffer, S. (ed.) (1950). *Measurement and Prediction.* Vol. IV of *Studies in Social Psychology in World War II*, Princeton, New Jersey: Princeton Univ. Press.
[36] *Technical Recommendations for Psychological Tests and Diagnostic Techniques* (1954). Washington, D.C.: American Psychological Association. (*Psychol. Bull.*, 1954, LI, supplement.)
[37] Tryon, R. C. (1957). Reliability and behavior domain validity: Reformulation and historical critique. *Psych. Bull.*, LIV, 229–249.
[38] Walker, D. A. (1931). Answer-pattern and score-scatter in tests and examinations. *Brit. J. Psychol.*, XXII, 73–86.
[39] Webster, H. (1960). A generalization of Kuder-Richardson reliability formula 21. *Educ. and Psychol. Meas.*, XX, 131–138.

Vol. XVI
Part 2 The British Journal of Statistical Psychology November
1963

ANALYTIC METHODS OF FACTOR ROTATION

By FRANK W. WARBURTON
Department of Education, University of Manchester

Recent work on analytical methods of factor rotation applied to large-scale data has emphasized the use of oblique factors and the simple structure criterion. The chief methodological advances have been (i) to use the variation of column (factor) loadings rather than row (test) loadings in order to reach a clearer simple structure (ii) the normalization of rows before rotation, in order to give the tests equal weight (this prevents complex tests, with high loadings in several factors, pulling the factors too close to each other) and (iii) the minimization of the covariance of the loadings (rather than the maximization of the variance) since this requires that high and low loadings occupy adjacent positions in the factor table in accordance with Thurstone's third condition for simple structure. The best method would appear to be Varimax in the orthogonal case, and Binormamin for oblique rotation. Maxplane is the closest to rotation by subjective methods.

I. INTRODUCTION

Work on analytical methods of factor rotation has drawn attention to several issues of considerable theoretical interest as well as of practical importance. In particular, it has stressed three main principles in factor analysis which have been increasingly observed in recent years, particularly in the U.S.A., namely, (i) the rotation of the factor axes away from their original bipolar positions, (ii) the use of oblique instead of orthogonal factors, and (iii) the use of simple structure as a criterion for rotation.

To illustrate these tendencies by an example. Post-graduate students taking a course in teacher training in the Education Department at Manchester University were given a test battery comprising (i) eight cognitive tests, (ii) the Kuder-Preference Record, (iii) the Allport-Vernon Scale of Values, (iv) Cattell's 16 P.F. personality questionnaire, and (v) Eysenck's inventory of social attitudes; the data also included (vi) 22 biographical particulars, e.g., age, interview grade, socio-economic status, and (vii) 16 criteria of success, either in practical teaching or in theoretical examination, making up a total of over 100 variables. The first striking feature of the resulting factor analysis is that 16 bipolar factors are extremely difficult to interpret, e.g., one has to pair intellectual ability + interest + stability *versus* socio-economic status + teaching ability + extraversion. Rotation becomes essential for purposes of interpretation, and one is forced to think in terms of 16 conceptually simpler rotated factors, which can be classified as follows: (i) non-verbal ability, verbal ability and verbal fluency in the cognitive field; (ii) the personality factors of stability and extraversion; (iii) the attitudes of radicalism and tender-mindedness; (iv) social factors such as family size, social status, social activities and interview assessment, and (v) the criteria of success represented by the factors of examination success and teaching ability.

Secondly, in large-scale analyses there are as many factors as there were tests in an old-fashioned battery, and the investigator finds it necessary, if the findings are to be satisfactorily interpreted, to have aknowledge of the inter-factor relations, i.e., to ascertain the correlations between the primary factors. Hence we need to resolve the individual tests into first order oblique factors (as distinct from orthogonal factors) with orthogonality introduced only at the second order level. The trouble with orthogonal methods applied to large-scale data is that the material is usually much more heterogeneous than was the case with the conventional old-time small battery, and that in consequence, with several dozen variables in the battery the orthogonal method tends simply to rotate the variables rigidly back into their original heterogeneous groups such as (i) academic attainment, (ii) social status, (iii) personality traits and (iv) moral values. This puts the research worker back where he started, and the old fallacy about only getting out of a factor analysis what you put in is, for once, true. In large-scale experiments it is therefore advisable to introduce another stage. We go from tests to first order primary oblique factors, and then from the primary oblique factors to second orthogonal factors. For example, in the Manchester research outlined above the primary factors grouped themselves into second order factors interpreted as Stability, Extraversion, Radicalism and Tendermindedness.

The third point brought into prominence by analytic rotation is that with large-scale heterogeneous data the criterion of simple structure becomes additionally attractive. One needs to know where a major influence is conspicuously present *and* where it is absent. No factor should load in everything. We cannot for example, satisfactorily define Spearman's '*g*' solely in terms of intellectual capacity unless we also know that it is absent in (say) measures of physical characteristics.

A fourth feature of the analysis of large-scale data is that *objective* methods of rotation become desirable, since it is impossible to rotate a hundred tests hundreds, or even thousands, of times by hand.

Let us look at some analytical methods of rotation in more detail. The general principle of rotation to a simple structure position is to obtain hyperplanes as large as possible, i.e., to rotate each pair of axes in turn to positions which will maximize the number of near-zero loadings. This is achieved by placing the primary axes through clusters of tests, taking care that the axes shall not be so oblique (within, say, 30° of each other) that they become, in effect, a single factor. In a four factor example, we may proceed by rotating successively factor 1 against factor 2, factor 1 against 3, 1 against 4, 2 against 4, and so on, round the complete cycle. Eventually the hyperplane count reaches a maximum which cannot be increased by any further rotations, i.e., we have attained a maximum number of zeros. Some research workers also aim at getting a maximum number of high loadings. An analysis which contains a large number of high and low loadings is said to approximate to simple structure, and it is very much easier to interpret than an unrotated bipolar analysis with its relatively large number of middle sized loadings. A typical example is given in Table 1 of the increasing hyperplane and high-loading count in a research on moral values in education (Nesbitt 1962).

TABLE 1. HYPERPLANE COUNT AND HIGH-LOADING COUNT WITH SUCCESSIVE ROTATIONS

Successive Rotations	% 'Low' Loadings (<0·10)	% 'High' Loadings (>0·49)	% 'Extreme' Loadings
1	43	10	53
2	43	11	54
3	42	12	54
4	48	12	60
5	51	14	65
6	52	16	58
7	61	16	77
8	60	14	74
9	64	14	78

Cattell (1960) has put forward a method of analytical rotation based on this procedure. It is called Maxplane, as its purpose is to maximize the hyperplane count. In this method factors are rotated in pairs. In Stage 1 the machine shifts one factor at a time, in five successive 9° shifts. It is not necessary to move beyond 45° in one direction, because the shifts on the other factor later explore downwards to the 45° position. Thus all 360° on the plot of these two factors are explored at 9° intervals through 40 shifts, 20 forwards and 20 backwards.

At each position of the axes the number of variables in the hyperplane of prescribed width, i.e., ± 0·10, is counted and the information stored. On the machine this requires that all shifts, fruitful or unfruitful, shall actually be completed, whereas the human rotator would evaluate the position by eye and decide not to proceed with less rewarding rotations.

A count is also made of the 'high' loadings, i.e., those above ± 0·49, since we wish the final evaluation to rest both on the goodness of the hyperplane count and on the variance of the factor—Cattell suggests that for most psychological studies the high-loading count will be between 10–50 per cent of the hyperplane count.

In stage 2 the process is carried out with shifts of 4° instead of 9°. It is common knowledge that the hyperplane as seen in any one plot will at first not be seen 'edgewise' but will consist of an ellipse. In Stage 2, however, the ellipse will have narrowed, by reason of the adjustments in other places, and it becomes appropriate to count now for a narrower band, say ± 0·05 instead of ± 0·10. Finally, in the third stage, the programme shifts to counting in that final, narrowest possible width which is fixed as four times the standard error ($\pm 2\sigma$) of a zero loading for the given study, i.e., so as to include 95 per cent of the true count. Reference vectors are prevented from approaching closer than 35° to each other, representing a maximum interfactor correlation of +0·82.

Convergence towards a maximum hyperplane count is not direct and continuous, but proceeds in a series of dampened oscillations in a curve rising to a plateau, but the general tendency is, of course, for smaller and smaller shifts of the axes to be made, and for the degree of accuracy demanded to be increased. The programme must have a built-in stopping point, analogous to the human decision that there is no point in going any further. At present the programme is made to stop when the current cycle fails to produce a 1 per cent increase in the hyperplane count. It seems, with psychological data, to be usually reached in about 12 to 36 shifts, which takes a reasonable amount of machine time. Maxplane's chief drawback so far has been that if a hyperplane crystallizes quite late in a rotation, at perhaps 30° or more away from where Maxplane is finally settling down, it will not (like a human computer) shift to an entirely new hyperplane 20° to 40° away from where it has hitherto been working. This happens fairly often in the clearing up process. The method would probably be improved by being programmed to make periodic wide (40°) sweeps at late stages in the procedure. In practice, Maxplane appears to be about as good as the more algebraic rotational methods, and not quite as good as a long, patient, highly skilled subjective rotation. It is too mechanical to yield quite as good a simple structure as a human being can achieve. Cattell found that about ten extra subjective visual shifts are usually necessary.

II. ORTHOGONAL ROTATION

It is clear that an algebraic method of attaining simple structure, based on maximizing the number of high and low loadings at the expense of the middle loadings, can be obtained by either (i) maximizing the variance of the loadings or (ii) minimizing the covariance of the loadings. But, since the variance of the rows is necessarily unchanged by rotation, the variance of the *squared* loadings has to be maximized.

The first effective formulae for analytical rotation came from Carroll (1953) although the methods were not applied until 1958 when electronic computing had sufficiently developed to make their application possible. Essentially the same theoretical basis—the maximization of the squared factor loadings—was developed in four separate researches by Carroll (1953), Ferguson (1954), Saunders (1954) and Neuhaus and Wrigley (1954). Each of these research workers regarded an approach towards simple structure as the best means of reducing the factorial complexity of the factor, and the four criteria led to identical results for the orthogonal solution.

Carroll (1953) considered Thurstone's (1947) five principles of simple structure, but he immediately ruled out the likelihood of a single mathematical expression embodying all these characteristics. His method, which he later named Quartimin, is described by Kaiser (1959) as " the first attempt to break away from an inflexible devotion to Thurstone's ambiguous, arbitrary and mathematically unmanageable qualitative rules for his intuitively compelling notion of simple structure ". Carroll considered the minimization of some sort of inner product function of the row loadings of the rotated matrix, and proposed that the sum of cross-products of the squared row loadings should be a minimum.

Ferguson (1954) used some of the concepts of information theory in arriving at his solution. He began by pointing out that the conditions for simple structure cannot be represented as terms in a single mathematical expression capable of manipulation. He approached the problem of parsimony by considering a single variable, represented by a point, and asking himself what was its most parsimonious description. He suggested that, intuitively, the most parsimonious description results when one of the axes passes through the point. It is seen that when the reference frame is rotated so that one of the axes approaches the point, the product of the two co-ordinates grows smaller. When one of the axes passes through the point it has a loading of zero on one factor and the sum of the products with any other factor is zero. Continuing this line of reasoning, Ferguson suggested that some function of the sum of the products of the co-ordinates for a set of collinear points might be used as a measure of the amount of parsimony attained in the description of these points. Finally, in order to deal with the problem of positive and negative co-ordinates he proposed as a measure of parsimony (in place of the simple sum) the sum of the squares of the products of the row co-ordinates.

Saunders (1953) suggested maximizing the kurtosis of the loadings since, as kurtosis is increased, the relative frequencies of the middle (near zero) loadings and tails (large loadings) of a normal distribution are increased at the expense of the intermediate regions. The statistical measure of kurtosis is the sum of the fourth powers of the loadings divided by the squared sum of squares, so that this criterion is basically equivalent to Carroll's.

Neuhaus and Wrigley (1954) suggested that the best criterion of simple structure would be the maximization of the variance of the squared factor loadings, i.e., the sum of fourth powers minus the squared sum of squares, hence it was called Quartimax. In the orthogonal case Neuhaus and Wrigley's function is Ferguson's criterion minus a constant, and Saunder's criterion is Ferguson's criterion divided by a constant. In the orthogonal case Carroll's function, too, is equivalent to the others. When the derivative of each of these is set equal to zero to solve for the maximum they give equal or proportional derivatives. Thus, four different functions developed according to four different theories prove to be equivalent mathematically.

Neuhaus and Wrigley subsequently used the Quartimax method in several investigations, but the results were rather disappointing. With experience, methodological advances have more recently been made in three directions.

(1) Emphasis has been placed on the variation of column (factor) loadings rather than row (test) loadings. This leads to a clearer simple structure, as the manipulation of loadings only by rows tends to perpetuate a general factor.

(2) The normalization of rows has been carried out before rotation begins, in an attempt to give the tests equal weights. (The rows are denormalized after

rotation is completed.) It is found that otherwise complex tests, with high loadings in several factors, play too large a part in the rotational process, and pull the factor axes towards each other, resulting in a high degree of obliquity.

(3) The minimization of the covariance of the loadings has taken precedence over the alternative method of maximizing the variance, since it not only demands a maximum number of high and low loadings but also requires that high loadings occupy adjacent positions to low loadings in the factor table, in accordance with Thurstone's third condition for simple structure, namely " for every pair of columns of the factor matrix there should be several variables whose entries vanish in one column but not in the other ", i.e., covariance methods (as distinct from variance methods) lay down conditions for the disposition of loadings as well as their magnitude. Minimization of the covariance has, in fact, proved a more effective method than maximization of the variance in the oblique case, although no difference is apparent for orthogonal rotations.

The difficulty with the Quartimax method, as Kaiser (1959) has pointed out, is that it works only by rows. It increases the larger values in each row and reduces the smaller values to near zero. This takes into account only the first of Thurstone's five rules for simple structure, namely, that each row of the factor matrix should have at least one zero. It is more important from the point of view of simple structure to increase the hyperplane count by reducing column rather than row elements, i.e., to aim at numerous zero loadings in each factor rather than in each test. This requirement precludes a general factor. Kaiser turned the Quartimax criterion on its side. He maximized the variance of the squared elements by columns instead of by rows. This is the ' Raw Varimax ' method. Nevertheless, systematic bias remained, owing to the divergent weights which implicitly are attached to the tests by their communalities: tests with high variance tend to swamp the rotational process, since when one deals with fourth power functions of the factor loadings, a test with a communality of 0·9, for example, will tend to influence the rotations nine times as much as a test whose communality is 0·3. Kaiser therefore substituted a ' Normal Varimax ' for the ' Raw Varimax ' criterion. Both, it should be emphasized, are based on maximizing the column variance, but Normal Varimax begins by giving the tests equal weight, the rows (i.e. tests) being normalized before rotation begins, and then denormalized after rotation is completed. Kaiser believes, however, that there is an even more important reason than the attainment of simple structure for establishing the Normal Varimax criterion as a mathematical model for the rotation problem. He found that the angle of rotation does not depend on the number of tests in each cluster, and subsequently that the Normal Varimax solution is invariant under changes in the test battery. This is not true for either the Quartimax or the Raw Varimax solution, since with these the angle of rotation is partly a function of the number of tests in the various clusters. Although the Normal Varimax solution was originally devised for the purpose of satisfying the simple structure criteria, Kaiser now claims that the attainment of simple structure is less of a fundamental advantage of his method than the attainment of factorial invariance. To secure numerical verification of the tendency of this criterion to give factorially invariant solutions Kaiser analysed Holzinger and Harman's battery of 24 psychological tests. A rotated solution was obtained for the first five tests, then a separate solution for the first six, and so on systematically, until the analysis for all 24 tests was reached. He found that some factors were essentially invariant from the outset; the changes in their loadings as the size of the battery was gradually increased were negligible. On the other hand, for the other factors the loadings showed more movement before they became stable, but this occurred only while they contained no appreciable variance. Once they picked up a test or two with high loadings their ambiguous definition stopped, and they settled down to a high degree of invariance. More recently Carroll (1959) has suggested that the best solution is given by Orthomax, a weighted combination of the Quartimax and Varimax methods.

The machine procedure for Varimax as programmed by Kaiser is fairly typical of analytical methods. It involves successive pairings of the m factors and carrying through the full $m(m-1)/2$ transformations for each cycle. The process is iterative until a complete cycle is obtained in which no rotation is required. When the machine finds the angle of rotation which will give a maximal sum of fourth powers of loadings of rotated factors it makes the transformation only if the angle of rotation is greater than some specified value, perhaps one degree, or only one minute if the computer has a very large high-speed memory capacity. Cycles of operation are continued until the sum of fourth powers for the entire matrix no longer increases beyond a certain limit. The convergence is generally rapid in the first few cycles and then quickly tends to slow down. We have also to decide the number of decimal points required. Usually the calculation of cycles is continued until convergence is obtained to the eighth decimal place. A typical example is the cycle of 78 separate rotations for Thurstone's primary mental abilities which was calculated on the Illiac in about $1\frac{1}{2}$ minutes, i.e., about one second per rotation for 57 tests. The sum of fourth powers converged to eight decimal places after 27 cycles—about 45 machine minutes in all, allowing for input and output time. More than 90 per cent of the increase in the sum of fourth powers occurred in the first four cycles, so that a reasonable solution might have been obtained in about ten minutes machine time. In practice, limitation to the fourth decimal place and one degree of rotation would appear to give as great a degree of accuracy as we need for practical purposes. Few factorists would venture to interpret the third (let alone the fourth) decimal place in a factor loading, or rotate through angles of less then one degree.

The computational procedures for considering factors two at a time do not necessarily give the best possible simple structure. The order in which the factors are paired may alter the results. Until a procedure is developed for operating on all factors at once, the method described here must suffice for an approximation.

III. OBLIQUE ROTATION

Analytical rotation in the orthogonal case would now appear to be on a sound theoretical and practical footing. On the other hand, fully satisfactory analytical methods for the oblique case have yet to be developed. In an orthogonal rotation the analytical groups are more or less equivalent, but unfortunately the row sums of squares are no longer constant in the oblique case, and the four methods are not equivalent. Only two in the orthogonal group—Carroll's original Quartimin criterion and Saunders' kurtosis criterion (called Oblimax in the oblique form)—have been extensively used for rotation to oblique simple structures, and all methods adopted (except Oblimax) are based on the minimization of cross-products, and are collectively termed Oblimin methods.

The objection to Quartimin, (in the oblique case), as Kaiser and Dickman (1959) have pointed out, is that the function is biased. Complex tests which load in a relatively large number of factors strongly influence the common positioning of the axes, and pull all the axes towards one another, yielding primary factors which are much too oblique, i.e., too highly correlated. They have found that the only case for which Quartimin operates satisfactorily is that in which every variable is loaded in one, and only one, factor, i.e., in the unifactor solution. On the other hand, the Varimax criterion which was developed specifically under the restriction of orthogonality, has been modified for the oblique case by Carroll (1957) and named Covarimin. It rests on the minimization of the covariance of squared loadings of columns after row normalization. He found that Covarimin also has a systematic bias, but in the opposite direction; it yields factor axes that are too orthogonal and gives factors of low intercorrelation. The Varimax criterion tends to maintain its orthogonality, even when modified to meet the oblique case by the Covarimin criterion.

Carroll noticed something else about the two badly biased criteria of Varimax and Covarimin for Thurstone's box problems. Their opposite and large biases were almost equal. He decided to split the difference and give the two methods equal weight, calling this the Biquartimin criterion. It was soon found, however, that the ideal split between the two methods is not necessarily an equal one; if the data are highly structured—nearly pure Covarimin is called for. As the structure becomes more complex, more and more Varimax is indicated, i.e., it counteracts the tendency for complex tests to have undue influence on the rotations. Later, Kaiser and Dickman (1959) found that column normalization helps to provide a parameter giving the appropriate proportion of Covarimin to Varimax. They have called this method Binormamin, since it is based on the two-way normalized matrix. Kaiser decided to extend the process of normalization by applying it to columns as well as rows. This equalizes the variance of the factors as well as the tests. This is the most recent, and on the whole the most successful, objective method of oblique factor rotation. All bias has not been removed by the use of Binormamin since it has a bias of the same sort, and in the same direction as Quartimin; the primary factors are too highly correlated. The bias in Binormamin is slight, however, compared with Quartimin, Covarimin or Oblimax. The bias in Binormamin is systematic and related to the complexity of the tests. For highly structured data there is almost no bias, but the bias increases as the items become more complex, i.e., related to more of the factors.

On the whole, the best method would appear to be Varimax in the orthogonal case, and Binormamin for oblique rotation. Maxplane is the closest to human subjective rotation. The various methods are summarized in Table I.

TABLE I. METHODS OF ANALYTICAL FACTOR ROTATION

Date	Authors	Name	Function	Remarks
ORTHOGONAL				
1953	Carroll	—	Minimization of the sum of products of the squared row loadings.	Called Quartimin in oblique form. Works only by rows and hence retains general factor.
1953	Saunders	—	Maximization of kurtosis = sum of the fourth powers of the loadings divided by the squared sum of squares.	Called Oblimax in oblique form.
1954	Ferguson	—	Maximization of the sum of the fourth powers of the loadings.	
1953	Neuhaus & Wrigley	Quartimax	Maximization of the variance of the squared row loadings = sum of the fourth powers minus the squared sum of squares.	Emphasizes a general factor.
1958	Kaiser	Raw Varimax	Maximization of the squared loadings of columns.	Emphasizes simple structure, but tests of high communality have too much weight.
1958	Kaiser	Normal Varimax	Maximization of the squared loadings of columns after normalization of rows.	Emphasizes simple structure.
1959	Carroll	Orthomax	—	Weighted combination of Quartimax and Varimax.
OBLIQUE				
1954	Saunders	Oblimax	Maximization of kurtosis = sum of fourth powers of loadings divided by the squared sum of squares.	Gives highly correlated factors. Less biased than Quartimin.
1957	Carroll	Covarimin	Minimization of covariance of squared loadings of columns after row normalization.	Inverse of Varimax. Gives low correlations between factors (like Varimax).
1957	Carroll	Quartimin	Minimization of the sum of the products of the squared row loadings.	Gives highly correlated factors.

1957	Carroll	Biquartimin	—	Combines Quartimin and Covarimin equally.
1959	Kaiser & Dickman	Binormamin	Minimization of the sum of the products of the two-way normalized matrix.	Weighted combination of Quartimin and Covarimin. Factors tend to be highly correlated, but less so than in Quartimin or Covarimin.

REFERENCES

CARROLL, J. B. (1953). An analytical solution for approximating simple structure in factor analysis. *Psychometrika*, XVIII, 23–28.

CARROLL, J. B. (1957). Biquartimin criterion for rotation to oblique simple structure in factor analysis. *Science*, CXXVI, 1114–1115.

CATTELL, R. B. (1960). The Maxplane program for factor rotation to oblique simple structure. *Ed. psych. Measmt.*, XX, 569–590.

DICKMAN, K. W. (1960). Factorial validity of a rating instrument. Ph.D. Thesis, University of Illinois.

FERGUSON, G. A. (1954). The concept of parsimony in factor analysis. *Psychometrika*, XIX, 281–290.

KAISER, H. F. (1956). Note on Carroll's analytic simple structure. *Psychotrmaeik*, XXI, 89–92.

KAISER, H. F. (1959). Computer program for Varimax rotation in factor analysis. *Ed. psych. Measmt.*, XIX, 413–420.

KAISER, H. F. (1958). The Varimax criterion for analytic rotation in factor analysis. *Psychometrika*, XXIII, 187–200.

KAISER, H. F., and DICKMAN, K. W. (1959). Analytic determination of common factors. Paper presented at the American Psychological Association Annual Convention, Cincinnati.

NEUHAUS, J. O., and WRIGLEY, C. (1954). The Quartimax method: an analytic approach to orthogonal simple structure. *Brit. J. stat. Psychol.*, VII, 81–91.

NESBITT, J. E. (1962). The development of character integration in children aged from nine to fourteen years. Ph.D. Thesis, Department of Education, University of Manchester.

PINZKA, C., and SAUNDERS, D. R. (1954). Analytic rotation to simple structure: II. Extension to an oblique solution. *Educational Testing Service Research Bulletin*, No. 54, 31.

SAUNDERS, D. R. (1953). An analytic method of rotation to orthogonal simple structure. *Educational Testing Service Research Bulletin*, No. 53, 10.

SAUNDERS, D. R. (1953). An analytic method of rotation to orthogonal simple structure. *Amer. Psychologist*, VIII, 428 (abstract).

THURSTONE, L. L. (1947). *Multiple Factor Analysis*. Chicago; University of Chicago Press.

Vol. XVI
Part 2
The British Journal of Statistical Psychology
November
1963

IS INTELLIGENCE DISTRIBUTED NORMALLY?

By Cyril Burt
University College, London

Frequency distributions obtained on applying intelligence tests to large samples of the school population are analysed, and compared with those given by the formulae for the commoner types of frequency curve. It is noted that the distributions actually observed are more asymmetrical and have longer tails than that described by the normal curve. The best fit is given by a curve of Type IV: this is in fact the type of distribution we should expect if (as has been argued in earlier papers) individual differences in general ability are determined partly by multifactorial and partly by unifactorial inheritance. It follows that the usual assumption of normality leads to a gross underestimate of the number of highly gifted individuals. The conclusions thus drawn are confirmed by a study of data from other sources; and various practical corollaries are deduced.

I. Problem

In recent controversies about the abilities both of schoolchildren and of adults a number of questions have repeatedly been raised, some theoretical, others eminently practical, which cannot be answered without some fairly precise knowledge of the way in which individual differences in such abilities are distributed. Hitherto most psychologists and educationists have assumed that the distribution of abilities conforms to the so-called 'normal curve of chance'. In the case of 'general intelligence' this alleged normality has often been cited as evidence in favour of some particular hypothesis about the nature or origin of the individual differences observed: on the other hand, several critics have maintained that the apparent normality is merely an artificial consequence of the way mental tests are standardized, and can therefore have no such implications.

Nevertheless, it has become increasingly clear that deductions based on the assumptions of normality may at times be highly questionable. In discussing what is called 'the pool of intelligence' educationists have varied widely in their estimates of the number of 'potential geniuses' available in the child population—potential geniuses being defined for such purposes as 'those with I.Q.s of 175 or upwards'. Not many months ago, in reply to a question put to him in Parliament, the Minister of Education gave an estimate of 'little more than one or two in a million'. The assessment, like others of its type, was obtained by employing the normal distribution with a conventional I.Q. scale and a standard deviation of 15 points. But, as the correspondence that followed quickly showed, the figure cited was considered much too low by many headmasters and educational psychologists, who found it quite out of keeping with the number of pupils with high I.Q.s who passed through their hands.[1]

The general acceptance of the theory that individual differences in ability are distributed in strict accordance with the Gaussian curve seems largely due to the advocacy of Thorndike, the acknowledged leader during the earlier decades of the century in the field of educational measurement. In his book, *The Measurement of Intelligence*, he devotes a chapter and a long appendix to demonstrating the conclusion that, providing the amount of intelligence is measured on a scale of 'truly equal units', the distribution should be exactly normal. On applying the chi-squared test to his own measurements he reaches values for P ranging from 0·99 to 0·999,999. Now we used to be warned that " a value of P very near to unity should lead the investigator to suspect his hypothesis quite as much as very small values: such very close correspondences are too good to be true " ([19], p. 423). But in any case the argument in practice tends to become circular: Thorndike's followers, at least in this country, were very prone to declare that, if the distribution did not conform to the normal curve, that showed that the units were not 'truly equal'.

Subsequent work in genetics has since furnished strong theoretical grounds for believing that innate mental abilities are not distributed in exact conformity with the normal curve. So far as they are inborn, individual differences in general intelligence are apparently due to a large number of genes of varying influence. Were inheritance solely 'multifactorial', i.e., if the genes consisted solely of numerous 'polygenes', each giving rise only to a very small deviation one way or another, then we might reasonably expect the resulting distribution to conform with the 'normal curve of chance'. But there can be little doubt that some of the genes are responsible for comparatively large deviations; and if their effects were sometimes favourable, sometimes unfavourable, the net result would be to enlarge the tails of the distribution in both an upward and a downward direction. However, since the genetic constitution of man is so delicately balanced and adjusted, the effects of these exceptional genes, or of the mutations that produce them, are more likely to be unfavourable than favourable. Hence, the final outcome will be a distribution that is more or less skewed, the longer tail being in the downward direction.

Since these exceptional genes are, by hypothesis, comparatively rare, it would seem to follow that, so long as we are concerned with small samples or with the general run of the population, a normal curve might still be trusted to yield a plausible fit. But when we are concerned with the more extreme type of deviation—the exceptionally bright and the exceptionally dull, cases which are so infrequent that they are only found in investigations covering very large groups—then the predictions deduced from the normal curve may easily prove mistaken. But—and this is a point which I wish most emphatically to stress—the nature of the distribution is not a matter to be decided (as is so commonly supposed) by mere 'assumptions', or even by deductive inference from general principles. It is an issue which can only be settled by an empirical investigation—that is to say, by an *ad hoc* analysis of data collected from actual surveys.

---

[1] See more especially *The Times Educational Supplement*, March 9 and 16, 1962. According to the report of the Parliamentary debates the estimate was said to have emanated from the National Foundation for Educational Research. In a personal communication, however, the Director, Dr. Wall, tells me that he has in point of fact never been " asked directly how many children there might be with I.Q.s over 175 ". He adds ," I myself would come to a very similar conclusion to the one stated in your paper: it is even possible that my estimate would be higher still ".

Oddly enough, nearly all of those who have joined in these discussions, whether as critics of the normality hypothesis or as supporters, seem to have missed the real reason for its popularity. It springs not so much from theoretical as from practical considerations. The ordinates and the areas of the curve were long ago calculated and tabulated once for all by Dr. Sheppard, later H.M. Inspector of Schools [15], and are now readily accessible in most popular textbooks on mental measurement. If on the other hand some other type of curve is assumed, the frequencies would have to be calculated *de novo* by each investigator for every fresh research. In the following analysis the chief novelty is the detailed comparison of the frequencies actually observed with theoretical values specially computed from the formulae for a curve of Type IV. Indeed, the primary object of the paper is not so much to supply better estimates for the number of children possessing this or that grade of intelligence, but rather to illustrate the practicability of more adequate methods of statistical analysis.

II. Data and Methods

The earliest of the statistical analyses which were carried out to gain light on the foregoing problems seemed plainly to indicate that the distributions of ability among school children by no means conformed to a strictly normal distribution (Burt, 1917, p. 34; esp. footnote 2, 1921, pp. 160f.). The two anomalous characteristics which might be expected to result from a combination of 'multifactorial' with 'unifactorial' inheritance—the elongated tails and the downward asymmetry—were already discernible in the frequency curves then obtained. The chi-squared test was regularly applied; and, wherever the samples were sufficiently large, the divergences from strict normality proved to be statistically significant. However, these initial surveys were merely experimental. The types of test used for such purposes—the original Binet–Simon scale and the earliest group tests—were predominantly verbal, and, as was to be anticipated, could claim no very high reliability or validity. On the whole and in this country the most efficient procedures now available for the purpose would appear to be the later British adaptations of the Stanford–Binet scales. Accordingly in what follows I shall confine myself mainly to data procured by this means.

In a previous publication [4] I have already reported results obtained during investigations undertaken to secure material for an English standardization of the original Stanford–Binet scale. It was then shown, not only that the discrepancies were statistically significant, but also that the values for the usual criteria (the so-called beta coefficients) indicated that the distribution should be regarded as belonging to Pearson's Type IV. Further results have since become available during investigations with the 'new revised Stanford–Binet tests' [16]. The total number of children assessed in the course of all these surveys amounts to 4,665. Their distribution is shown in the first column of Table I. It is plainly skewed, with a prolonged lower tail. Of the entire group more than 10 per cent have I.Q.s under 80; only 7·7 per cent have I.Q.s over 120.

TABLE I. OBSERVED AND THEORETICAL DISTRIBUTIONS

Frequencies in Percentages

I.Q.	Observed		Theoretical		
	Before Screening	After Screening	Normal	Type IV	Type VII
Below 30	0·11	0·02	—	0·03	0·01
30–	0·06	0·02	—	0·03	0·01
35–	0·09	0·04	—	0·05	0·03
40–	0·21	0·11	0·01	0·09	0·05
45–	0·23	0·16	0·03	0·14	0·09
50–	0·39	0·27	0·10	0·27	0·17
55–	0·62	0·46	0·26	0·43	0·33
60–	0·77	0·64	0·62	0·74	0·62
65–	1·35	1·24	1·32	1·30	1·71
70–	2·34	2·08	2·54	2·28	2·16
75–	3·97	3·40	4·37	3·71	3·77
80–	6·41	5·70	6·75	5·86	6·16
85–	9·01	9·07	9·37	8·78	9·18
90–	11·60	11·96	11·61	11·47	12·13
95–	13·01	13·42	12·94	13·79	14·02
100–	13·42	13·84	13·01	14·11	14·09
105–	12·67	13·07	11·66	12·57	12·18
110–	10·01	10·33	9·41	9·66	9·22
115–	6·00	6·19	6·76	6·38	6·18
120–	3·45	3·56	4·38	3·92	3·79
125–	1·97	2·03	2·54	2·15	2·16
130–	1·20	1·24	1·32	1·12	1·17
135–	0·60	0·62	0·62	0·57	0·62
140–	0·28	0·29	0·26	0·28	0·33
145–	0·11	0·11	0·10	0·13	0·17
150–	0·06	0·07	0·03	0·06	0·09
155–	0·04	0·04	0·01	0·04	0·05
Above 160	0·02	0·02	—	0·04	0·05
Total	100·00	100·00	100·00	100·00	100·00

However, with each of the component batches we have endeavoured to eliminate all those cases in which there was the smallest reason to believe that the low I.Q. was mainly or largely due to non-genetic causes, either environmental, pathological, or accidental (e.g. caused by injury at birth).[1] The number remaining[2] after these eliminations were made was 4,523. The data for this composite group have now been analysed by Miss Baker along the lines which I adopted in reporting the separate surveys. To conform to the conventional scale now in current use all the I.Q.s have been restandardized so as to yield a standard deviation of 15 points.[3]

---

[1] One critic asks for a brief explanation of our use of the word 'pathological'. In previous reports we followed Dr. Lewis's terminology (adopted in the Wood Report), and used the word to designate those cases that are apparently due to non-genetic or post-conceptual disturbances. Some writers, however, have included under the term cases showing clinical symptoms that result, not only from environmental causes, but also from rare recessive or dominant genes. For this mixed group Penrose's term 'clinical' seems more suitable ([11], pp. 46f.). His data and that of other investigators suggest that something like three-quarters of the certifiable defectives are of a mixed 'clinical' type; the other quarter—the so-called 'residual' cases—may be regarded as almost wholly genetic. Among the 'clinical' cases the defect in about one-third of the total number seems attributable, partly at any rate, to genetic causes. The rest form the group we have called 'pathological'. However, the real difficulty is to classify those cases—perhaps the most numerous of all—in which both types of factor operate and in which environmental factors tend to aggravate and often to obscure factors that are essentially genetic. Sometimes it is possible to allow for the joint causation by raising the observed I.Q. by 10 to 20 points; but any such classification (as was emphasized above) is bound to be somewhat speculative. In our own assessments we have sought to err on the side of eliminating too many rather than too few.

[2] This total is made up of three main batches. (i) A set of 433 children tested during the course of a preliminary study of the Stanford–Binet scale: for these, in the original L.C.C. report, 1925, Terman's age-allocations were adopted, but the I.Q.s have now been restandardized. (ii) A set of 2,835 children, tested in the course of a joint research planned to secure data for an English re-standardization of the Stanford–Binet scale ([3], p. 348). The distribution of these first two batches combined into a single group, is shown in [9], p. 56 and Figure 1. (iii) The latest batch consists of 1,255 children tested with the 'new revised tests', i.e. the so-called Terman–Merrill scale. The restandardization of the Terman–Merrill revision was undertaken by a small committee originally formed under the chairmanship of Professor Hamley, with Miss Baker as secretary. We should like once again to express our indebtedness to all those who assisted in the work or allowed us to carry out tests in their schools.

The shortcomings of such a composite sample are obvious. In the London surveys those who co-operated in the testing usually had received a training in such work, and were given personal instructions; the results could be referred to the children's teachers, and where necessary checked by further study or by different types of test; in particular, the majority of the subnormal children in the area belonging to the age-group (including defectives not attending school) were tested and included in the survey. With the surveys carried out in other areas all this was seldom possible. In these the number of subnormal children is probably disproportionately small, especially as we were anxious that if errors were made, they should be made on the safe side—i.e. by excluding too many genetic cases and not by including non-genetic cases: otherwise it might be argued that the gravest cases of subnormality were pathological.

[3] We have kept to measurements expressed in terms of the I.Q. in order that the results may be intelligible to the general reader and comparable with those obtained in other investigations. Such a unit is not without its drawbacks. For example, if we reintroduced all the cases of subnormality, and re-calculated the standard deviation, it would turn out to be much larger. Some may argue that this is the standard deviation we ought to take as our unit, since it is based on the entire population. But in another educational area the entire population would probably yield yet another value. And in any case most investigations exclude the more extreme cases of mental deficiency from their calculations. As I have argued elsewhere the whole problem of a suitable mental unit calls urgently for reconsideration; but that is a side-issue which cannot be examined here and does not affect our present problem.

The distribution of the group, after this preliminary screening had been carried out, is shown in the second column of Table I. In I.Q. points the mean is almost exactly 100, and the standard deviation almost exactly 15 (calculated to two decimal places they are 100·07 and 15·09 respectively). These values of course are the results of the standardization. The amount of individual variation, however, differs in the two different directions: in the upper half of the curve it corresponds to a standard deviation of only 14·8, and in the lower half to one of 15·4.

Column 3 in the table shows the theoretical distribution calculated from the usual tables for the normal curve and based on the calculated mean and standard deviation given above. It is at once obvious that the frequencies found in the survey rise in the centre to too sharp a peak and exhibit tails that are far too widely and unequally spread out for the observed distribution to be regarded as typically normal. On applying the usual test for goodness of fit, we find $\chi^2 = 107{\cdot}3$, and consequently, if the distribution in the general population were strictly normal, it would be well over a million to one against discrepancies so large as those observed occurring as a result of the mere chances of random samples. At the same time it will be noted that, with a much smaller sample (e.g., one which only justified us in calculating the percentages to one decimal place and contained no individuals with an I.Q. over 150 or under 60—as indeed is the case with many of the frequency tables published for surveys), the normal curve would provide a very plausible fit. (See Figure 1, which shows

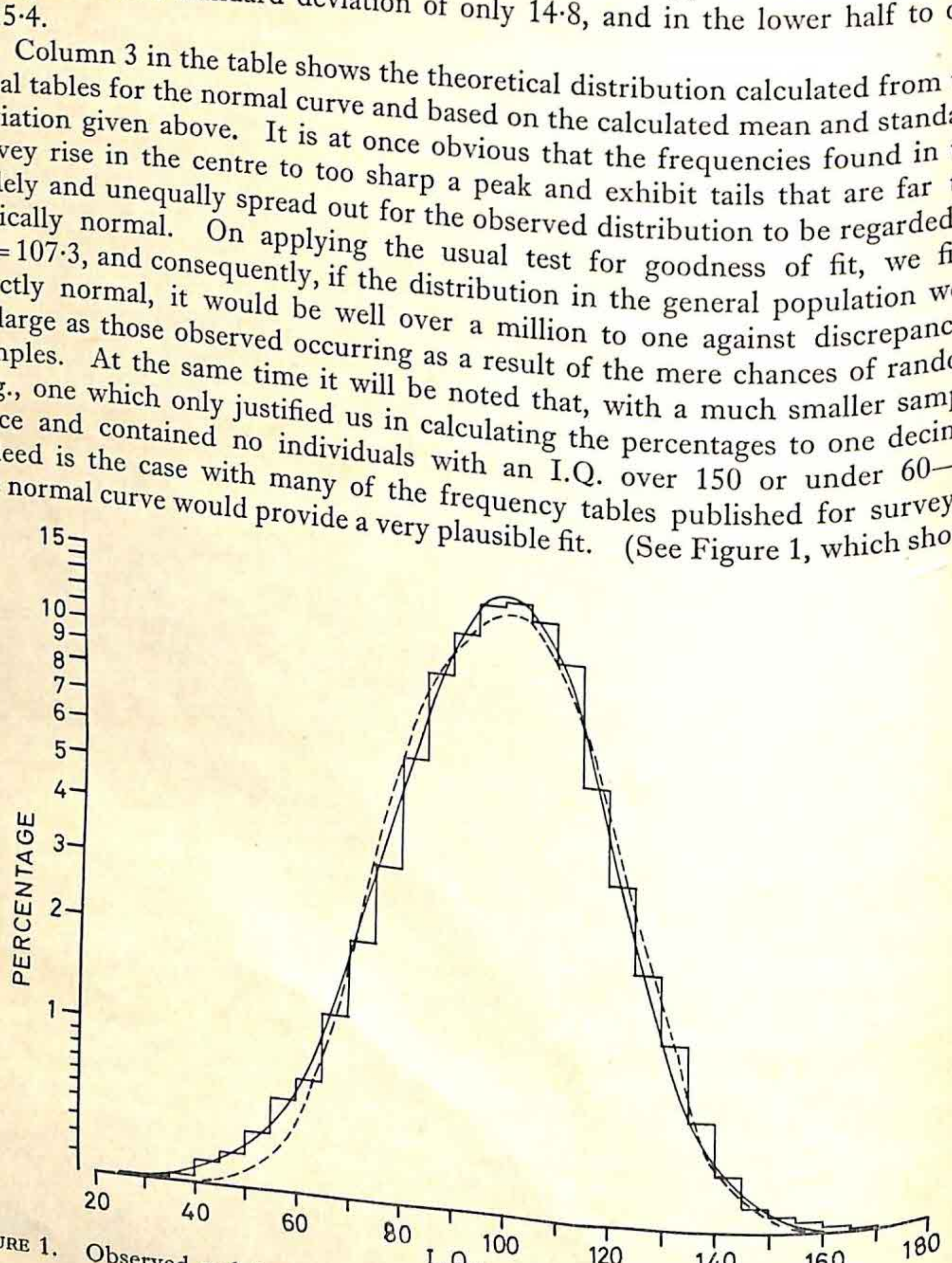


FIGURE 1. Observed and theoretical distributions. (Type IV continuous line; Normal Curve dotted line.)

the observed frequencies after the doubtful cases have been eliminated by screening, and the best-fitting curves of the ' normal ' type and of type IV.)

In order to discover to what class of curve the distribution belongs, we have here followed the same procedure as before. Using the notation suggested by Karl Pearson we find β_1 (the criterion for skewness) $=0{\cdot}048$, and β_2 (the criterion for kurtosis) $=3{\cdot}918$. This suggests that the curve is either Type VII or Type IV. Taking Fisher's forms of the criteria, we obtain $g_1=-0{\cdot}219\pm0{\cdot}036$ and $g_2=0{\cdot}918\pm0{\cdot}073$, and from these values or from the betas we obtain[1] $\kappa=0{\cdot}007$. Curves for which the criterion κ lies between 0 and 1 are classified by Pearson as belonging to Type IV. This is the only type in his series which is (i) of unlimited range in both directions and at the same time (ii) asymmetrical and (iii) leptokurtic, i.e., peaked.

To determine how closely a hypothetical curve of this type will fit the actual data it will be necessary to compute the theoretical frequencies. The method adopted is virtually that which I described and used in my previous reports, and is based (with minor modifications) on the second of the two working procedures discussed by Palin Elderton ([6], pp. 67f). Instead of the simpler formula which he uses we have preferred the slightly more complicated expression in which the deviates to which the frequencies refer are deviates from the mean instead of from the arbitrary origin, namely,

$$y=y_0\left\{1+\left(\frac{x}{a}-\frac{\nu}{r}\right)^2\right\}^{-m}\exp\left\{-\nu\tan^{-1}(x/a-\nu/r)\right\} \tag{1}$$

where y_0 is the ordinate at the mean, and r, m, ν, and a are constants computed from β_1 and β_2, and a varies with the standard deviation. With the present data the values are $r=10{\cdot}184$, $m=6{\cdot}092$, $\nu=1{\cdot}527$, and $a=9{\cdot}501$. This yields values for the ordinates at mid-points of the successive intervals into which the total range is subdivided. To effect a valid comparison we need areas rather than ordinates; and for this purpose we have applied the formula

$$\int_{-\frac{1}{2}}^{+\frac{1}{2}} y_x\,dx=\frac{1}{24}\left(y_{x-1}+22y_x+y_{x+1}\right) \tag{2}$$

where y_x denotes the ordinate at the mid-point of the interval and y_{x-1} and y_{x+1} the ordinates next below and next above. Elderton observes that the calculation of the values relating to this curve " needs considerable care: it is ", he adds, " the most difficult of all the Pearson-type curves ". We have found the labour lengthy rather than difficult; but admittedly there are ample opportunities for slips and mistakes.

---

[1] The small size of κ is due to the small degree of asymmetry combined with the high degree of lepto-kurtosis. In the London sample, where special efforts were made to locate all the low-grade cases (including defectives not in attendance at a school) the values for β_1 (0·035) and κ (0·04) were appreciably larger, and the asymmetry was far more conspicuous. Readers unfamiliar with Pearson's criteria and his scheme for classifying frequency curves may refer to the *Note* on the subject recently published in this *Journal* (XV, 1962, pp. 80f., especially Figure 1, p. 86).

III. Results

Let us now glance at the results and conclusions reached on adopting these somewhat novel formulae. The detailed frequencies computed by the foregoing equations are shown in the fourth column of Table I. χ^2 is now only 18·2; $P=0{\cdot}57$. The fit is perhaps not quite so close as that obtained by Karl Pearson for distributions derived from physical measurements. But it is clear that the discrepancies between the theoretical values and the values observed may now quite well be due to the chances of sampling, and that the agreement is far better than that commonly obtained in cases of mental measurement.

To estimate the proportional number of individuals having I.Q.s above any borderline which lies outside the range of Table I—e.g., the borderline of 175 I.Q. mentioned in the question put to the Minister of Education—we must undertake some kind of extrapolation. With the aid of the formula we can extend the theoretical calculations beyond the limits reached by our sample and sum the areas. Table II shows the chief results. Here the figures in column ii represent the cumulative frequencies thus obtained, i.e., the proportional numbers reaching or exceeding the I.Q. specified in the left hand margin. Column i gives the proportions deduced from Pearson's tables for the normal curve, assuming as before a standard deviation of approximately 15 I.Q. It will be seen that according to our estimates the number having I.Q.s of over 160 is more than ten times the number deduced from the normal curve, and that, instead of the proportion of those with I.Q.s of 175 or over being only 3 or 4 per million, it rises to nearly 77 per million. At the present day the male school population in England and Wales amounts to rather over 4 millions. That would yield more than 300 boys of school age with I.Q.s over 175. Among the female population, just as the number of defectives is smaller than that obtaining among the males, so apparently is the number of geniuses. But in any case, the total number of children reaching the high level specified cannot be far short of five or six hundred.

Table II. The Estimated Proportions of the Population Reaching or Exceeding the Borderline Specified

Borderline	Number per million	
	(i) Normal Distribution	(ii) Type IV Distribution
160	31·7	342·3
175	3·3	76·8
190	0·1	19·4
200	<0·001	6·2

It is however of interest to consider whether a simpler formula might not furnish a reasonable fit. We have seen that, when pathological cases are eliminated, the degree of asymmetry becomes comparatively slight; both β_1 and κ have decidedly low values. Accordingly let us ask what type of curve would give the closest fit if we assumed that the apparent asymmetry resulted solely from errors of sampling. This would imply that the true value of β_1 and κ is zero. In that case the constant ν, which is based on β_1, would likewise be zero. The exponential factor in eqn. (1) then becomes unity, and the equation reduces to the extremely simple form

$$y/y_0 = \left(1 + \frac{x^2}{a^2}\right)^{-m} \tag{3}$$

—the formula for a curve of Type VII. The equations for the two constants are correspondingly simplified. Substituting the numerical values for β and σ (the standard deviation) we now obtain

$$a^2 = \frac{2\beta_2}{\beta_2 - 3}\,\sigma^2 = 77{\cdot}876 \tag{4}$$

and

$$m = \frac{5\beta_2 - 9}{2(\beta_2 - 3)} = 5{\cdot}770. \tag{5}$$

With the aid of a table of logarithms the theoretical frequencies can be readily computed. The figures obtained are shown in column 5 of Table I. Applying the usual test we have $\chi^2 = 23$, and $P = 0{\cdot}28$. The fit is decidedly better than that given by the normal curve. If therefore the distribution in the general population was in fact a distribution of Type VII, the probability that discrepancies as large as this or larger would occur as a result of random sampling would be just under 1 in 3.

The well-known relation between the very simple formula thus reached (eqn. 3) and the more familiar formula for the normal curve is worth a passing comment for the benefit of training college lecturers and others concerned with elementary courses on statistical psychology. The relation turns on the fact that, as n increases indefinitely, the limit of $(1 + 1/n)^n$ is the number e, the base of the natural logarithms, and the limit of $(1 - z/n)^n$ is e^{-z}. Now let us write eqn. (4) in the form $a^2 = -\sigma^2 k^2$, then eqn. (5) gives $m = -\frac{1}{4}k^2\ (5 - 9/\beta_2)$; so that, as β_2 approaches 3, m approaches $-\frac{1}{2}k^2$, and k^2 increases indefinitely. In that case eqn. (3) will take the form

$$y/y_0 = \left(1 - \frac{x^2}{\sigma^2 k^2}\right)^{\frac{1}{2}k^2},$$

or, putting $z = x^2/2\sigma^2$ and $n = \frac{1}{2}k^2$

$$y/y_0 = \mathrm{e}^{-x^2/2\sigma^2}$$

when k^2 becomes indefinitely large.

Moreover, it will be noted that, if in eqn. (3) we put $a^2 = n$ and $m = \frac{1}{2}(n + 1)$, we obtain the equation which expresses the distribution of Student's t. Suppose we take m as approximately 6; then n will be 11; and, on referring to the table for t given in Yule and Kendall ([19], p. 537), it will be seen that with a standard deviation of $t = 11$, the values in the column for $n = 11$ yield on subtraction proportionate frequencies which are not unlike those in column 4 of Table I, and so offer a rough fit to data such as the present.

IV. Supplementary Evidence

Critics who still wish to defend the hypothesis of strict normality may perhaps be inclined to suggest that, since our list of observed frequencies has been obtained from a composite sample, the two peculiarities we have noted—the asymmetry and the elongated tails—might well be just the incidental consequences of the ways in which the constituent groups have been selected and combined. In reply may I point out that each of the constituent groups themselves showed the same features, and in particular, with the largest group of all—based on a survey in which we sought to include the entire child population between the ages of 6·0 and 11·0 residing in a representative electoral division of London—the calculated values for the beta coefficients indicated, even more clearly, a curve of Type IV (cf. [4], p. 170)?

It is instructive to compare the results obtained in these and later surveys with those found by the American investigators while standardizing the latest revision of the Stanford–Binet tests, i.e., the so-called Terman–Merrill scale ([10], pp. 21–23). For this purpose a group of just under three thousand children was tested with the two alternative forms of the scale—L and M. With form L, the I.Q. ranged from 35 to 170, with form M from 35 to 165. The average of the whole group, however, was appreciably higher than ours—104·0 I.Q. with form L and 104·4 with form M. This means that the range runs from 70 points below the average to 60 or 65 points above—limits which are not very different from our own. On applying the chi-squared test for agreement with normality, McNemar found, for form L, $P=0{\cdot}03$ and for form M, $P=0{\cdot}005$. Thus, particularly in the latter case, the divergences from the normal curve are fully significant. The discrepancies of the fit, though numerically large, are by no means obvious to the eye when inspecting the various graphs that show for each scale the observed frequencies superimposed on the best-fitting normal curve: (*op. cit.*, figs. 1 and 2, p. 19). The relevant constants are as follows. For form L, $g_1=0{\cdot}028\pm0{\cdot}045$, and for form M, $0{\cdot}029\pm0{\cdot}045$; for form L, g_2 is $0{\cdot}346\pm0{\cdot}090$ and for form M, $0{\cdot}298\pm0{\cdot}090$. Thus both frequency distributions are significantly leptokurtic as the diagrams certainly suggest on a closer examination: both seem too sharply peaked to be regarded as normal, and the frequencies at the end of either tail extend too far. On the other hand, the criterion of skewness is in both cases non-significant. However, as McNemar points out, there were several unavoidable defects in the sampling. The distribution, we are told ([17], pp. 15f.), contained an unduly small proportion of children from the lowest occupational groups: indeed, to judge by the table (*loc. cit.*, p. 14) the percentage in the lower group (unskilled day-labourers) was apparently only one-third of what it should have been. Moreover, since the sample was restricted to schools—and indeed to what are described as 'average schools'—and therefore did not include subnormal children living at home or in institutions, the number of defective as well as of dull and backward pupils must have been disproportionately small. Had they been included, the downward asymmetry, which appears in 4 out of the 6 constituent groups, would be much more clearly marked. McNemar himself makes no attempt to identify the type of distribution shown. He expressly refrains from drawing any conclusions from these data concerning the probable distribution of intelligence, and merely observes that " the I.Q.s are *approximately* normal in distribution ". And of course, so long as we are dealing with the ordinary run of children in the primary and secondary schools, the assumption is not likely to lead us far astray.[1]

However, scattered throughout the relevant literature there is a good deal of additional evidence which strongly suggests that for those who are definitely subnormal the relative frequencies deduced from the normal curve are far too low. In his discussion of the " medical grouping of institutional cases " Penrose, for instance ([11], p. 45), points out that " far too many individuals exist, whose abilities are more than 3 or 4 times the standard deviation below the normal mean, to be fitted under a Gaussian curve: on that assumption only about 1 idiot among 10,000 and 1 imbecile among 6 could belong to a normal population with a standard deviation of 15 I.Q. points ": with his classification, as his table shows, the ' observed percentages ' are 0·04 and 0·24 respectively, whereas a normal distribution would predict only 0·000,004 and 0·06. Unfortunately in official reports on mental deficiency the figures commonly given have been collected for administrative purposes, and relate only to those defectives who have been formally ' ascertained '. Nearly always they are expressed as fractions of the total population not of the relevant age groups; nor do they, as a rule, attempt to distinguish between the cases that are undoubtedly pathological and those that are presumed to be wholly or partly of genetic origin. Moreover, owing to the variation in the standard deviation and the inadequacy of the tests employed, the I.Q.s used in defining the border-lines are by no means equivalent to the I.Q.s on the conventional scale that is now in use among psychologists.[2] When a reasonable allowance is made for these disturbing factors, the estimates for the incidence of subnormality, as shown in nearly all the published surveys, both British and American, nearly always suggest proportional numbers far higher than those which would be deduced from a distribution that was strictly normal (see [5] and refs.)

---

[1] Unfortunately the data from the two Scottish surveys throw little light on the problem. In the first survey the group tests produced " a very skewed distribution "; but this the investigators ascribe to the " ceiling of the test " which tended to curtail the upper end ([14], p. 108); for the Binet tests no special schools were included, and consequently the frequency curve drops abruptly at about 70 I.Q. ([14], pp. 91f.). In the second survey special schools and institutions were visited, but " testing was undertaken only when the pupils' handicaps would not invalidate the test " ([15], pp. 9 and 57): with the Terman–Merrill tests, which were applied to over 1,300 pupils, more than 10 per cent had I.Q.s above 160 and nearly 1 per cent I.Q.s of 170 or more. Thus the figures obtained lend no support to the assumption of normality. But it could be argued that quite possibly much of the discrepancy was due to the peculiarities of the tests and samples rather than to the actual distribution of the general population.

[2] Let me repeat that those who think of comparing the frequencies reported in this section with those obtained in their own inquiries should bear in mind that the I.Q.s referred to are not the raw quotients calculated by dividing mental age by chronological age—a method which yields standard deviations anywhere between 13 and 17 I.Q. points or more according to the set of tests employed. As explained above, the performances of the children tested were first re-scaled—a point which critics of my previous papers seem to have overlooked. And yet, as we have seen, even in the present inquiry the standard deviation of the *composite* group turned out to be a fraction over that 15 points. Thus I.Q.s of 70 and 130 represent not ±2 S.D. exactly, but only ±1·988 S.D.; and this is the value adopted in deducing the theoretical frequencies given in the text.

This point, however, has been sufficiently stressed in earlier publications. The exact determination of assessments for supernormal ability has received far less consideration from previous investigators. In recent discussions about the 'national pool of ability' and the 'need for more adequate methods of selecting and training the undiscovered reserves of ability' attention has been chiefly concentrated, both in this journal and in the popular press, on the numbers at the higher end of the scale—more especially the number of potential entrants to the grammar schools and to the universities, i.e., of those whose abilities rise above levels of about 110 and 130 I.Q. Here I should like to insert a special plea for a more adequate recognition of the needs and the numbers of those I have called the '*exceptionally* gifted'. They form a group who constitute one of the nation's most valuable assets, and whose special educational requirements have hitherto been grossly neglected. The average I.Q. of pupils with I.Q.s over 160 (about 170 I.Q.) must be as far above the average I.Q. of the general mass of grammar school children (about 122 I.Q.) as the average of these selected children is above that of the educationally subnormal (i.e., pupils with I.Q.s below 85, whose average would be about 77 I.Q.). Such highly gifted individuals must therefore feel as much out of place in the ordinary grammar school as the grammar pupils would in a class of dull and backward youngsters; and this is fully borne out by evidence gathered from such youngsters while still at school, or later on when, as adults, they have reported or recorded their school experiences.

Let us therefore glance first of all at one of the very few factual inquiries relating to these exceptionally gifted children. In Terman's investigation the initial aim was to select and study those whose I.Q.s would place them "well within the top 1 per cent of the school population". The working borderline was fixed at 140 I.Q. Of those selected the number with I.Q.s of 170 or over amounted to 6·7 per cent, i.e., between 300 and 600 per million in the general population according as we interpret "well within 1 per cent". The proportion, as several reviewers pointed out at the time, was unexpectedly high; and perhaps some allowance should be made for the fact that the average level of the population in the cities of California was said to be 4 or 5 points above the general average of the U.S. population, while the standard deviation (with Terman's age-allocations) was nearer 16 points than 15 ([17], pp. 19, 44f). In a later search for exceptionally gifted pupils in New York City—a search described as by no means 'exhaustive'—Dr. Leta Hollingworth found at least a dozen with I.Q.s over 180, i.e. a proportion of about 20 per million. Deviations of this size, she tells us, would be expected "only once in more than a million times if the distribution corresponds to Quetelet's curve of probability"; but, she adds, "it seems more likely from existing data, that children who test above 180 I.Q. are present in greater frequency" ([8], pp. xiii, 23f.). Still more recently the Counseling Centre of New York University reported a follow-up study of a batch of children who at the time of testing (approximately age 5) had I.Q.s of 170 or more; and in that inquiry over a hundred cases were discovered.

In this country the figures I and my co-workers have obtained in the course of various surveys have varied widely from one area to another. In my earliest inquiries, carried out with the assistance of the Department for the Training of Teachers at Oxford, I found, in a single age-group (aged 9½ to 10½) numbering approximately 1,600 children in all, six children with test-scores equivalent to an I.Q. of 175 or above, nearly 0·4 per cent; nearly all of them came from one particular preparatory school which catered specially for the children of dons. On the other hand, in Liverpool, within a group of the same age but nearly eight times as large and consisting solely of pupils attending public elementary schools, not a single child of this high level was discovered; outside the elementary schools, however, a number of such cases were located, and we estimated that their proportional frequency amounted to just under 30 per million. In my first survey of Council schools in London I found 3·24 and 0·45 per cent with deviations exceeding twice and three times the standard deviation respectively, i.e., with I.Q.s of 130 and 145, instead of 2·27 and 0·13 per cent as we should expect with a strictly normal distribution ([2], pp. 161, 174f.; 4th ed., pp. 199, 218f.); the London child whom we discovered with an I.Q. of 190 (the case cited by Dr. Hollingworth in the discussion just mentioned) was encountered in a private school. However, at the time of the earliest L.C.C. surveys comparatively few parents belonging to the professional classes sent their children to the public elementary schools, and the test then used—which consisted either of those forming the original Binet–Simon scale or those adopted in our first attempts at group testing—scarcely did justice to children who were exceptionally bright. At a later date, when carrying out inquiries for the Consultative Committee of the Board of Education, we found at what were then called ' secondary schools ' 23 boys between the ages of 11 and 16 with I.Q.s of 175 or over—most of them attending ' Headmasters' Conference Schools '. But the parents of two of these ' exceptionally gifted ' children lived outside London, and three others had come to London specifically for the sake of the child's education, which reduces the number of genuine Londoners to 18. In the County of London the male population between those ages then amounted to just over 200,000 which would suggest a proportion of about 90 per million.[1]

Finally, for the benefit of readers who may feel doubtful about a purely statistical approach we may try an alternative method of estimation by following the lines adopted by Galton in his early study of genius ([7], p. 34). With this aim in view let us consider a single generation—that of men born in the

[1] Since the above was written, Mr. G. C. Robb, Educational Psychologist for the Lincoln and Lindsey Education Authorities, tells me that, out of 1,085 children who took the 11-plus examination in Lincoln City last year, 29 scored an I.Q. over 130 with the Moray House Test. On examining them individually with the Terman–Merrill scale (Form L) he found 7 who scored more than 170 I.Q. However, Terman himself gives the standard deviation at 11 and 12 or 18 and 20 I.Q. points. Hence on the scale we ourselves have used Mr. Robb's figure of 170 would at this age probably be equivalent to about 160 or less.

British Isles during the first 30 years of the 19th century. I choose this period rather than any other, first because with any later period it would be hard to compile an agreed list of the most eminent persons, and secondly because during an earlier period it is highly likely that many geniuses of humble origin, such as Faraday and Dickens, might have failed to develop or manifest their true powers; nor would it be easy to secure satisfactory evidence for assessing their I.Q.s.

Since few eminent people succeed in demonstrating their claim to the title of genius before they reach middle life (say 45), we must confine ourselves to those males who survived to at least that age; and for simplicity of calculation let us keep to round figures. The total number in the generation selected would be about two and a half millions. Now in the case of five men of eminence born within the period chosen—John Stuart Mill, Sir William Rowan Hamilton, Lord Macaulay, Lord Kelvin, and Sir Francis Galton—we have detailed records of their early childhood sufficient to indicate that their I.Q.s must have been approximately 200. This is already a proportion of about 2 per million. But many other equally famous names will spring to mind—men born within the same dates who reached the same level of eminence, but for whom the records of childhood are less informative. Looking through the *Dictionary of National Biography*, we find that, roughly speaking, there were well over a dozen in each of the following categories: (1) men of letters (poets, novelists, essayists, historians, etc.), (2) classical scholars and philosophers, (3) mathematicians and scientists, (4) men of eminence in industry or commerce, and rather less than a dozen (5) politicians and lawyers, and (6) men of eminence in other fields—engineers, architects, painters, musicians, explorers, military commanders, etc. This makes about 75 in all out of 2,500,000, or almost 30 per million. We are, however, frequently assured, that, owing to the handicaps of poverty and social class, there must have been in the 'under-privileged' groups quite as many 'mute inglorious Miltons' who had no opportunity to develop their latent abilities, and so died with all their music in them. To make allowance for these unknown abortive geniuses we ought at least to double the figures. We should then reach a proportion of about 60 per million.

However, a critic may object that we have hit on an unusually productive era.[1] I myself should be inclined to agree that, owing to the diminished birth-rate among the professional and upper classes and the mortality during the first world war, the proportion of those with I.Q.s of 200 or more would today very probably be smaller. Accordingly let us meet this criticism by lowering the borderline rather than the per-millionage. We might still reasonably maintain that the proportion with I.Q.s over 175 would reach, and in all likelihood exceed, the figures just cited. Thus the estimate reached by these broader considerations would seem to be quite in keeping with that which we deduced from our Type IV curve—viz. somewhere between 60 and 80 per million.

---

[1] Galton's own estimate of the proportional number of 'geniuses' at the time when he was writing (1865) was 250 per million: indeed, he adopted this figure as his definition of genius. However, on scrutinizing the names in his list, it seems clear that only a small proportion of these could have had I.Q.s of over 180, and many probably had I.Q.s nearer 160. He himself assumed a normal distribution. But a careful study of his data and his analyses shows clearly that the number of those with I.Q.s over 175 is far above the figure we should deduce from a normal distribution.

V. Summary and Conclusions

1. A detailed analysis of test results obtained from a large sample of English children (4,665 in all), supplemented by a study of the meagre data already available, demonstrates beyond reasonable doubt that the distribution of individual differences in general intelligence by no means conforms with strict exactitude to the so-called normal curve. In the present inquiry an endeavour has been made to exclude all cases where non-genetic factors appeared to play an important part; and, in doing so, in order to forestall any suspicion as to a preconceived bias, we have always sought, wherever any doubt arose, to deal quite ruthlessly with any cases which might tend to impair the requirements of a normal distribution. When this is done, the discrepancies remaining are admittedly by no means large. Nevertheless, they appear to be fully significant statistically.

2. The main divergences are due to an elongation of the tails in both directions and to a marked tendency towards negative asymmetry. These are in fact the divergences we should expect if, as I have argued elsewhere, unifactorial as well as multifactorial modes of inheritance are the basic causes of the individual differences so found. The calculated constants suggest that the distribution actually obtained can be fitted much more appropriately with a curve belonging to Pearson's Type IV. The discrepancies between the actual frequencies and the theoretical frequencies calculated on this assumption prove to be statistically non-significant. Type IV is the type we should expect from the combined operation of the two modes of inheritance; it is also the type that yields the closest fit to the majority of distributions reported for biological data generally.

3. For many practical purposes, particularly when dealing with small samples and with children falling within a fairly limited range, the tabulated values for the normal curve will give a reasonable approximation. But they furnish very misleading assessments for the number of individuals showing an extreme deviation in either direction from the general mean. The fact that the number of *subnormal* individuals is far larger than that which would be expected from a strictly Gaussian distribution has long been suspected by various writers, both on genetic and on statistical grounds. What seems hitherto to have escaped general notice is the fact that the number of *supernormal* individuals also shows a similar though slightly smaller excess. And the evidence here summarized indicates that the current assumption of normality has led to a gross underestimate of the number of highly gifted children in the school population. Thus the proportion of those with I.Q.s over 160 proves to be more than 12 times as that deduced from the normal curve, and the proportion of those with I.Q.s over 175, instead of being only about 3 or 4 per million (assumed in several official or semi-official statements), must be at least 70 per million, probably more.

4. The data on which we have chiefly relied were for the most part obtained from investigations concerned with the standardization of the tests employed; and it is always somewhat precarious to answer questions about one set of problems from information collected primarily to answer a different set. For the more general conclusions the evidence appears fairly convincing. But the detailed figures, both for the statistical constants and for the various frequencies, may possibly need some degree of modification. In particular, for the reasons stated in paragraph 1 above it seems likely that the figures here reached may actually underrate the discrepancies between the actual distribution and that which would be inferred from the assumption of strict normality.

5. In view of the practical importance of the issues thus raised it is to be hoped that one or more of the larger educational authorities will in the near future plan and carry out a systematic survey with the express purpose of securing more exact and trustworthy specifications.

References

[1] Burt, C. (1917). *The Distribution and Relations of Educational Abilities.* London: P. S. King.
[2] Burt, C. (1921). *Mental and Scholastic Tests.* London: P. S. King. 4th ed. (1962), Staples Press.
[3] Burt, C. (1935). *The Subnormal Mind.* London: Oxford Univ. Press.
[4] Burt, C. (1957). The distribution of intelligence. *Brit. J. Psychol.*, XLVIII, 161–175.
[5] Clarke, A. M. and A. D. B. (1958). *Mental Deficiency.* London: Methuen.
[6] Elderton, W. P. (1938). *Frequency Curves and Correlation.* Cambridge: At the University Press.
[7] Galton, F. (1869). *Hereditary Genius.* London: Macmillan.
[8] Hollingworth, L. S. (1942). *Children Above* 180 *I.Q.* New York: World Book Company.
[9] Mayer-Gross, W., Slater, E., and Roth, M. (1954). *Clinical Psychiatry.* London: Cassell.
[10] McNemar, Q. (1942). *The Revision of the Stanford–Binet Scale.* New York: Houghton Mifflin.
[11] Penrose, L. S. (1954). *The Biology of Mental Defect.* London: Sidgwick & Jackson.
[12] *Report of the Mental Deficiency Committee* (1929). London: H.M. Stationary Office.
[13] Scottish Council for Research in Education (1933). *The Intelligence of Scottish Children.* London: University of London Press.
[14] Scottish Council for Research in Education (1949). *The Trend of Scottish Intelligence.* London: University of London Press.
[15] Sheppard, W. F. (1903). New Tables of the probability integral. *Biometrika*, II, 174–190.
[16] Terman, L. M. (1925). *Mental and Physical Traits of a Thousand Gifted Children.* Stanford: Stanford. Univ. Press.
[17] Terman, L. M. and Merrill, M. A. (1937). *Measuring Intelligence.* London: Harrap.
[18] Thorndike, E. L. *et al.* (1927). *Measurement of Intelligence.* New York: Teachers' College, Columbia University.
[19] Yule, G. U. and Kendall, M. G. (1937). *An Introduction to the Theory of Statistics.* London: Griffin.

Vol. XVI Part 2 The British Journal of Statistical Psychology November 1963

THE INTERACTION OF HEREDITARY AND ENVIRONMENTAL INFLUENCES[1]

By Raymond B. Cattell
University of Illinois

This paper is concerned with improvements in the multiple abstract variance analysis method (MAVA). The assumptions of the latter are more clearly developed and implications for genetic, psychological, and sociological research issues are discussed. The main conclusion is that although it offers solutions to questions not answerable by twin study methods, or by current experimental designs related to learning theory, to personality theory and to important sociological issues, MAVA demands much larger and more carefully planned and highly organized research programmes than have hitherto been contemplated in psychology.

I. Introduction

Recent papers by Burt [1952], Burt and Howard [1956, 1957], Anastasi [1959], Fuller and Thompson [1960], Vandenberg [1961], Isaacs [1962], Loehlin [1962] and others show that the relative importance of environmental and hereditary influences has become once more a controversial topic in psychology. But it has become so in a new sense. For whereas the tendentious debates of thirty years ago begot more heat than light, the technical attention of today is generating methodological and conceptual improvements, and there is a growing awareness that advances in fields so diverse as learning theory and personality theory have been impeded by a bottleneck in knowledge, due to half a century of neglect of psychological genetics.

The central problem in evaluating the inheritance of mental traits has always been that in *psychological* measurement, hereditary and environmental influences seem inextricably intermingled. In many physical traits, like those of Mendel's peas, the characters of the genotype can be seen in the phenotype. But the direct analysis of *behavioural* inheritance by Mendelian unifactorial principles has only rarely been achieved, e.g. in Huntington's chorea (Spillane and Phillips [1937]), colour blindness (Burt [1912, 1946]); and perhaps a few disorders such as schizophrenia (Kallman [1946]). Elsewhere, psychologists have had to turn to the alternative of estimating the genetically and environmentally determined contributions to observed variances in a given population (Burt and Howard [1956] p. 126, assuming a purely multifactorial inheritance). From such nature-nurture ratios, a further attempt may be made to infer the nature of the polygenic mechanisms.

A decade ago the writer proposed two radical improvements in determining these ratios [1953]. The first consisted in substituting the measurement of unitary factors, in place of single, arbitrary variables. The second was to abandon dependence on twin study, and utilize a new method called the Multiple

---

[1] This investigation was supported by Public Health Service Research Grant MH 01733-17 from the National Institutes of Mental Health.

Abstract Variance Analysis Method [1953, 1955, 1960]. This (MAVA) resembles a method developed by geneticists in animal husbandry. However, it would be a mistake to consider them one and the same because in psychology this development was not only historically independent but conceptually different. It differs because we cannot control independent genetic and environmental influences in human beings, and because the latter produce social as well as biological families, whereas cows do not.

Some knowledge of the main MAVA design will here be assumed, and only a brief sketch given. Essentially it derives certain abstract or hypothetical variances, such as between family hereditary variance and within family environmental variance contributions from a set of simultaneous equations. In each of the latter some actually observable variance, such as that between sibs (taken in pairs) reared apart, unrelated children reared together, etc., is equated to the appropriate abstract, hypothetical sources and their covariance. For example, the observed variance for sibs normally raised together is written:

$$\sigma^2_{ST} = \sigma^2_{wh} + \sigma^2_{we} + 2r_{wh.we}\sigma_{wh}\sigma_{we} \qquad (1)$$

where σ^2_{ST} is observed variance in sib pairs reared together (cf. Burt and Howard [1956], eqn. 34): that is to say, the mean of two sib scores on the given personality factor is subtracted from the score of each. The squares of the two equal deviation values are then summed, over many families, and divided by appropriate degrees of freedom to give the value on the left hand side of (1). The abstracted or inferred variances are σ^2_{we} which is the 'within-family environmental' variance, and σ^2_{wh} which is the within-family hereditary variance. The term $r_{we.wh}$ is the correlation of these sources of variance. Several other equations like (1) can be set up, for sib pairs whose members are reared apart, unrelated foster children reared together, and other diverse combinations of hereditary and environmental contributions [Cattell, 1955, 1960]. From these simultaneous equations each having on the left a value empirically determinable from a particular family constellation one can solve for the unknown desired 'abstract' contributory variances. Compared with the MAVA method, the only other design used by psychologists—the twin method (e.g. Burt [1961], p. 6, eqn. ii)—is relatively powerless. For it yields no information about the correlation of hereditary and environmental influences and it gives a ratio only for the peculiar environmental difference between twins (which numerous psychologists and psychoanalysts have shown to be different from that in typical sib families).

Among the criticisms which the MAVA design, like any new method, has rightly had to face, are three which deserve serious consideration. First, it is objected that the formulae sometime assume that certain values are identical which in fact might be different and therefore should be represented by two unknowns instead of one. MAVA has been worked out (Cattell, 1960) in a full design, with eleven unknowns, and also a limited resources design with only ten equations, and the latter assumes the correlation between hereditary and environmental influences to be the same for fraternal twins as for ordinary brothers. Such initial economy of assumption is necessary because the number of unknowns for which we wish to solve—basically four unknown variances and six unknown correlations—treads very closely on the number of independent equations we can hope to set up for getting empirical observations.

Here, as in any armchair formulation of possible scientific concepts it needs no great skill in casuistry to assert ever finer possible differences in one's shades of meaning for this or that kind of environment. What matters is whether these definitions correspond to any real classes, and whether too much hair splitting may not finally defeat the scientific aim of an effective economy of explanatory concepts. Of course, the ideal procedure is to *enter* with as many terms as one can conceive to be necessary, even erring on the side of entertaining a few luxury assumptions, and to leave the empirical verdict to decide whether what one has treated as a variable can rightly be regarded as a (superfluous) constant. Actually, a whole series of degrees of economy have been reviewed for the MAVA model, ranging from eight to twelve unknowns [Cattell, 1960]. The latter involves getting equations and data from half-sibs and identical twins reared apart—difficult to get in sufficient numbers—but even when reduced to finding eight unknowns the MAVA method is dealing with vastly more refined conceptions and greater comprehensiveness of solution than the classical twin method.

A second objection to the method has been that children in foster families are not biologically typical of the nation's children as a whole. Perhaps the onus of proof for this should be returned to the critic. But, if we assume he is right, one answer would be to match our control 'normal' families, on sociological parameters for origin, status, etc., with those containing the sibs reared apart and the unrelated children reared together, in order to bring about simultaneously some match in biological range. Parenthetically, let us note that this suggestion of selection of some kind in the variance of the adopted child is not the same as the assertion that social placement agencies place children in such a way as to bring about nature-nurture correlations. The latter is not likely to amount to anything except in the case of intelligence, in regard to which it is asserted that some placement agencies choose the educational level of the adoptive home to fit the level of intelligence of the child, as expected from his biological parentage. Though non-random placement is unlikely except in intelligence, it is catered for in the existing MAVA method, which introduces and solves for a possibly non-zero placement correlation. Another way of avoiding the objection that foster children are a biologically abnormal sample would be to drop those equations and observations which deal with children raised outside their own biological families. This could be done; but only at the cost of laming the design by cutting it down to a relatively coarse set of equations and unknowns.

The third difficulty is that the standard errors of the estimates for the within and between family hereditary and environmental variances, etc., as obtained by the MAVA method are terrifyingly large. This requires a fuller discussion and is taken up in Section V.

II. The Environmental-Genetic Relation as Covariance and Interaction

A normal sequence of exposition might seem to require that we discuss the meaning of the sources of variance before discussing their interaction. But the former, having the usual meaning and mode of calculation, present no mysteries for the average reader which cannot be postponed to later stages in the discussion (Section V). The correlations and interactions of heredity and environment, however, have some unexpected and intriguing features, the clarification of which is essential to a primary survey of the model.

The simplest MAVA model begins with the assumption that hereditary and environmental influences are additive and linearly related to their effects. Unlike the twin method it does not assume they are uncorrelated. Both the linearity and the correlation are illustrated by eqn. (1) above for sibs reared together. The various penetrating discussions of the problem, especially the considerations in the contributions of Wright [1960], Fisher [1930], Burt and Howard [1956, 1957], Penrose [1959] and more recent writers on genetics make it extremely probable that significant correlations of hereditary and environmental influences will exist both within and between families.[1] However, *a priori* analysis is now supported by evidence. The first uses of MAVA [Cattell, Blewett and Beloff, 1955: Cattell, Kristy and Stice, 1957] furnished a crop of significant correlations, both in the within and the between family situations. Among the latter were an r of $+0{\cdot}25$ between heredity and environmental effects on intelligence, between families (remarkably close to the independent value[1] commonly obtained for family intelligence and social status), and an r of $-0{\cdot}56$ (also between families) between hereditary and environmental contributions to personality factor U.I. 23 [Cattell, 1957]—a factor which Eysenck [1953] calls neuroticism and which we have called ergic regression [Hundleby, Pawlik and Cattell, In Press]. These newly discovered correlations are not only of sociological interest, but have contributed firm ground to personality theory. The negative correlation with ergic regression suggests a self-correcting tendency in which families of greater impulsivity develop a tradition of greater restriction on emotionality. Indeed, the interesting and unforeseen finding that the majority of such correlations are negative is one of the two main supports for the *law of coercion to the bio-social mean* [Cattell, 1957].

Although the model thus respects the reality of correlation or covariance it has not so far admitted what is statistically known as 'interaction'. The latter implies that two or more independent variables are related to a dependent variable by something more than additive and linear relations: 'Independent' here means independent in conception and action, not statistically independent, or orthogonal: for we have already agreed that hereditary and environmental influences are correlated. The formula for scores on the variables, corresponding to the non-interactive relations of the variances, as expressed in (1) above would be:

---

[1] A volume of opinion to the contrary exists, however, among those less professionally immersed in the problem. For example, Davis, Eells and Tyler [1951] in American, and Isaacs [1962] in Britain assert there is no reason to believe that native intelligence is in any way correlated with social status.

$$d_{ST} = k_{wh}h_w + k_{we}e_w \tag{2}$$

where d_{ST} is the observed deviation from the mean between sibs reared together, and e_w and h_w are the standard (deviation) scores of environmental and hereditary influences within the family. (Subscripts for the individuals, i, are omitted; and instead of including the k's as hidden in the score scaling, as in eqn. (1), we here consider both environment and heredity to be measured in standard scores and the magnitudes of their relative influences to be shown by k_{we} and k_{wh}.)

If now we assume that interaction as well as covariance exists, any kind of product of e_w and h_w, raised to any powers, could be added to (2). In this *embarras de richesse*, where only some *one* specific formulation can be tested, it is reasonable to take the simplest, and to add only a product term, $e_w h_w$, weighted by a coefficient k_{hew}, which is independent of k_{we} and k_{wh}. This could be interpreted, for example, as meaning, in the case of intelligence, that a good environment operates more powerfully to improve intelligence when the native intelligence is already high. The resulting analysis into abstract variances would be (for any within family relation, i.e. omitting now the specificity of sibs together, ST),

$$\begin{aligned}\sigma^2{}_{ow} = {} & k^2{}_{wh} + k^2{}_{we} + 2r_{wh.we}k_{wh}k_{we} + k^2{}_{hew}\sigma^2{}_{hew} \\ & + 2r_{we.hew}k_{we}k_{hew}\sigma_{hew} + 2r_{wh.hew}k_{wh}k_{hew}\sigma_{hew}\end{aligned} \tag{3}$$

(ow means ' observed within '; the σ_{ew} and σ_{hw} terms vanish because they are in standard scores).

It might at first appear that we have complicated our solutions by introducing four new unknowns; but $\sigma^2{}_{hew}$, $r_{wh\cdot hew}$ and $r_{we.hew}$ can be derived from σ_{hw}, σ_{ew} and $r_{wh.we}$, so that only one need be introduced, namely, k_{hew}, the weight given to interaction. Consequently in the equations for all family constellations only two unknowns would be added; k_{hew} for within family interaction, and k_{heb} for between family interaction effects.

The prospects for employing a MAVA solution for an interaction as well as a covariance model are therefore reasonably good. It would mean either finding two new, independent simultaneous equations from new family constellations, e.g., half sibs reared apart, or dropping two of the correlations discussed in the next section on grounds that they must be very small, or adopting the approximation that interaction within and between families are much the same, thus reducing k_{hew} and k_{heb} to k_{he}. Actually, as discussed elsewhere, the hope of finding new family constellations to permit equations solving for more than the present twelve unknowns is reasonably good. For example, one could add (a) adopted and non-adopted sib pairs in the same family, and (b) adopted pairs from near relatives (nephews, second cousins), Furthermore, it must always be remembered that the MAVA can be applied to other types of groups than the family in which hereditary and social differences are variously combined. An independent evaluation of the MAVA design in terms of necessary groups, etc., has recently been made by Loehlin (In Press).

III. THE POSSIBLE VARIETIES OF HEREDITY-ENVIRONMENT CORRELATION

Having committed our model to heredity-environment correlation it behooves us toask to what these correlations may be due, in sociological, biological and psychological terms. Such an examination will show that a correlation term may be omitted in some family constellations, whereas in others a correlation appears which would not be expected when applying traditional rules of analysis of variance. Also, such discussion prepares the way for enlightened interpretation of any significant values experimentally detected.

In an earlier discussion of these correlations [Cattell, 1955; Cattell, Blewett and Beloff, 1959] we discussed the usual expectation that among four variances there could be six correlations. These would consist of four involving relations of within and between family deviations and two—$r_{bh.be}$ and $r_{wh.we}$—confined to terms both of which are either between or within. It was pointed out that the usual expectation from analysis of variance design practice that the first four would vanish because of the assumption of non-correlation of any individual's within group variance with his family's between group variance no longer holds. This, the first reason for solving for extra correlations beyond $r_{wh.we}$ and $r_{bh.be}$, rests on the fact that we are dealing with abstract variances, not concrete variances. It is a mathematical impossibility in groups consisting of random samples to get within and between variance significantly correlated. But while this is true of the total concrete variance, it is still possible for within-group hereditary variance to be correlated with between-group environment, or environment in the former with heredity in the latter, for these are special abstractions from the concrete whole.

The question therefore is no longer whether within-between correlation is possible, but whether it is likely enough to justify entering a non-zero term to get a numerical solution. Any psychologist can think of several theories which would require correlation *within* the family, as when parents toss more intellectually demanding problems to their more gifted child. A demonstrated instance is the $r_{wh.we}$ of $-0{\cdot}50$ which we found [1959] for the 'dominance' factor, E in the 16 PF., suggesting that sibs behave encouragingly towards the less dominant and fix 'trash for overtopping' (in Hamlet's terms) upon the more dominant. Similarly, any sociologist can generate hypotheses requiring family differences in heredity to become associated systematically with differences in environment to yield a significant $r_{bh.be}$. Although brighter children will tend to be brought up in the atmosphere of bright parents, one must also recognize the directional effect in the tendency of bright children to have been brought up in the environment provided by less bright parents and dull children by less dull parents than themselves, owing to regression to the mean.

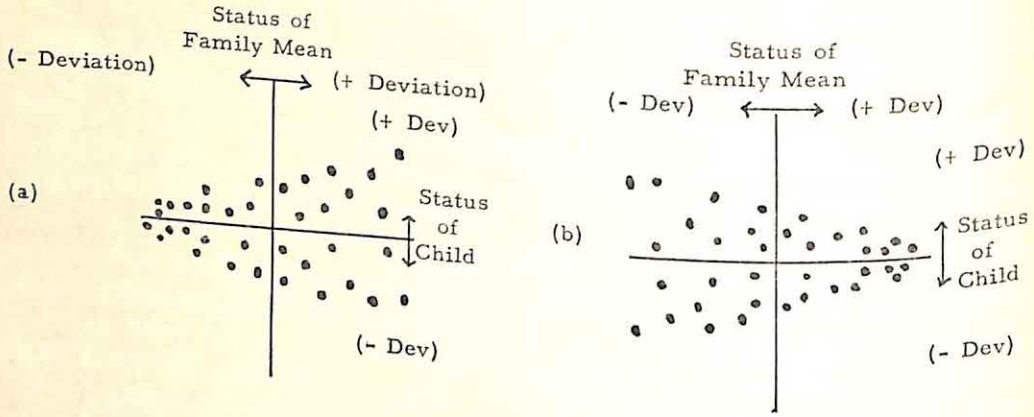
But one must take a more sophisticated look to see instances of within and between correlations. An instance of $r_{wh.be}$—an individual's likelihood of greater genetic deviation within his family being related to greater (or lesser) deviation of the environment of his family from the mean of families—is provided by the fact that genetically less assortative matings will tend to yield greater within family hereditary variance in the children. Since differences of the parents are part of the common family environment of the children, a child more deviant from his family hereditary mean is likely to be brought up in a family in which the environmental effects of having more 'discordant' parents are greater. A systematic relation in the converse direction, to lead to a significant, $r_{bh.we}$, would be illustrated by the theory that families with a high manic-depressive genetic endowment will tend to create an environment more uneven for different children (depending on the age at which they happen to experience parental instability).

The 'homogeneous' relations of within-environment with between-environment, $r_{we.be}$, and of within-heredity with between-heredity $r_{wh.bh}$, would not, mathematically, be able to yield any linear correlation, for the usual reason in anovar. They could, however, yield relations of the kind shown in Figure 1. In (*a*) for example, the families which deviate positively in environment could systematically contain individuals who deviate more in either direction than do families below normal. E.g. a family high in interest and capacity for giving educational aid to its children might discriminate among them more, according to capacities, in the education given, whereas in a family with low educational interests a general mediocrity of educational achievement might prevail. The opposite direction of association, equal in 1 (*b*), could be illustrated by the tendency of parents with greater and more enlightened affection for their children to bestow it more evenly, whereas careless favouritism might prevail in parents less conscientiously interested in their children. An instance of another pattern of relation of within deviation to between deviation in heredity (Fig. 1 (*c*)) could result if matings are less assortative at lower than at higher levels. (The data, Table 2 in Higgins, Reed and Reed [1962], suggests this may be true in the case of intelligence.) Then greater within-family hereditary variance would prevail at lower hereditary family levels.

Effects of the kind of (*a*), (*b*) and (*c*) could yield significant curvilinear correlation of absolute parent and child scores, in etas; and in (*d*) could result in a correlation coefficient never before described which could be formulated as 'double-eta'. However, none would yield a covariance term; and we are therefore left with the need to enter MAVA solution with four correlations only—$r_{wh.we}$, $r_{bh.be}$ $r_{bh.we}$ and $r_{wh.be}$.

At this point one has to take note of two, possibly three, peculiar psychological correlations. First, as already noted, when children are adopted there may be a tendency for agencies to place children of certain prospective maturations in families environmentally selected to fit those expectations, e.g., the child of a dead genius to the adoptive family of higher education. [There is little documentation of how far this goes on, consciously or unconsciously, but it is generally believed to be a serious possibility.] Accordingly we have introduced into the MAVA design the principle of *placement correlations*, indicated by a double prime, thus $r''_{wh.be}$ represents the correlation *within* families, and $r''_{bh.be}$, the correlation *between* families, produced by the above mechanism. It may be questioned whether the former (which implies that placement agencies can distinguish in the first years of life between sibs to go to a better and to poorer homes respectively) can be seriously entertained, but we have played for safety.

In Environmental Influences:



In Hereditary Influences:

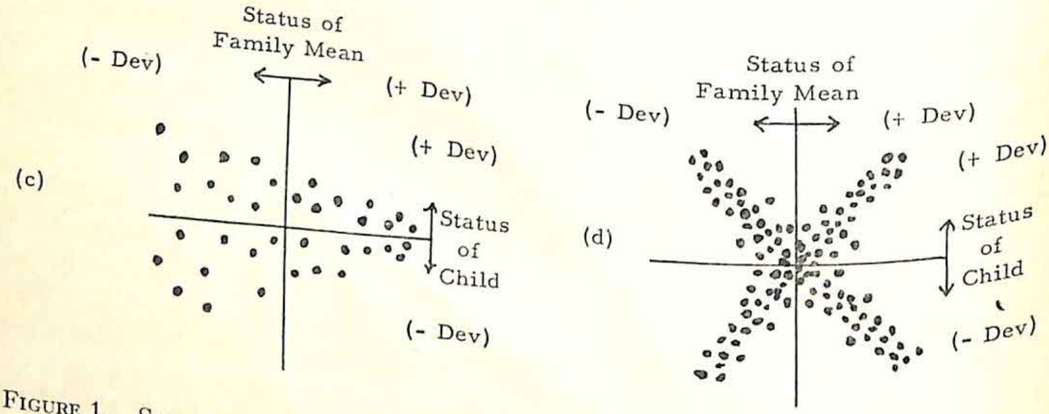


FIGURE 1. Some major examples of forms of association of within and between family deviations (and, therefore, variances) which do not create correlations. Scattergrams to illustrate hypothetical relations produced in variances. Substantive examples of these are discussed in the text.

Another class of correlations is that created by any need to change a correlation because of a change in range. For example, we have agreed in the standard MAVA formulas to a $\sigma^2_{we'}$ which accepts the presumed fact that differences of environment between twins are different from those between ordinary sibs—σ^2_{we}. (This raises the unknown variances to five). The question now arises whether the value $r_{wh.we'}$ is likely to be different from $r_{wh.we}$. *Against* this it might be urged that although $\sigma^2_{we'}$ is probably smaller than σ^2_{we} yet the *rate* of change of environmental treatment with hereditary difference would tend to be the same.

A similar problem arises in adoptive families where the range of hereditary variance is greater because it include the normal within family variance plus that due to the children coming from different biological families, thus $\sigma^2_{UT_h} = \sigma^2_{wh} + \sigma^2_{bh}$ (where *UT* means unrelated together). Should we here anticipate new values, $r'_{wh.we}$ differing from $r_{wh,we}$ because of the greater range.[1]

IV. EVALUATIONS OF ORIGINS AND POSSIBLE SIGNIFICANCES OF CORRELATIONS

Although we have discussed variance relations systematically, through all possibilities of the *model*, we have handled them only sporadically, and for purposes of illustration only, in terms of psychological, sociological and biological causes. Both to prepare ourselves for the interpretation of any correlations found to be significant, and to design experiment with a proper priority in sacrificing assumptions, it behooves us to consider systematically the real origins of correlation.

Broadly, one has to consider three possibilities, (1) that a genetic difference brings about an environmental difference, (2) that an environmental differential brings about a genetic association, and (3) that some third influence produces both. Since heredity and environment are supplementary and exclusive the third possibility may seem illogical; but it must be remembered that the environment which affects the child is not synonymous with the universe.

1. GENETIC CAUSATION OF ENVIRONMENTAL ASSOCIATIONS

Parenthetically, one must beware of confusing any true genetic effects on environment with the fallacious concept sometimes appearing in such statements as 'the same environment is more stimulating of intelligence in the intelligent child because he sees more in it'. This way lies madness! The only clear handling of perception is that which ascribes part of any perception to the meaning of the situation, given by its physical definition and the nature of the species [Cattell, 1962] and part to the individual. The former is what we are defining as independent environment here and anything else will show up in interaction terms.

---

[1] Earlier [1960] we have also split $r'_{wh.we}$ into an $r'_{bh.we}$ term. The latter seems unnecessary if we accept r'_{whwe} to hold steadily over the actual within family hereditary range due to both normal within family hereditary variance and to that component coming from between family hereditary differences. This is less confusing because if $r_{bh.we}$ enters in any other equations it does so in quite a different sense from this $r'_{bh.we}$, being due to effects of the kind discussed and exemplified above by, say, more intelligent families giving a more varied environment.

(*a*) *Within Family*

(i) *Associations by Migration.* The genetic endowment causes the individual to move into a certain kind of environment, as when a schizothyme temperament causes a person to take up such occupations as forestry or research physics, where experience in social situations is reduced.

(ii) *Associations by Shaping Environment.* The genetic endowment results in environmental correlation because it changes environment (as birds come to experience a nest environment, denied to animals). Countless examples can be cited. A fairly subtle instance is the finding by Cattell, Blewett and Beloff [1957] of the apparent tendency of a high I factor (tenderminded) person to create an environment which cherishes aesthetic development ($r_{we.wh} = +0{\cdot}10$ to $+0{\cdot}50$).

(iii) *Associations by Producing Environmental Response.* From the social and the physical environment a genetic endowment may provoke characteristic environmental treatment. The examples found by Blewett, Cattell and Beloff [1955] of negative $r_{we.wh}$ for dominance, suggests such a process. Here must be included the tendency of other people than the subject to select their associates according to preferred characteristics.

(*b*) *Between Family.* All three of the above operate here also, and an example combining them is the upward social mobility of families endowed with high intelligence. One would assume that correlations from (iii) would be smaller here than within families, because society does not know each family as well as a family knows its members. Conversely (i) and (ii) should be larger, because there might be less error in the mean performance of a group than in the fate of a single individual.

(iv) *Associations by Differential Survival.* If a certain genetic endowment favours survival by differences in birth rate or death rate within a certain social environment then $r_{bh.be}$ correlations will ensue.

2. ENVIRONMENTAL CAUSATION OF HEREDITARY ASSOCIATIONS

(*a*) *Within Family.* In general correlation within the family from this cause would be zero, since by hypothesis the child is already born. However, differential death rates might operate, e.g. an overprotective attitude towards the child might favour his survival if he is of weak constitution, increasing the conjunction of overprotection with defective constitution.

(*b*) *Between Families.* All four mechanisms under 1 can act here and in the reverse direction—environment as cause. By natural selection environment, stimulating migration, could produce correlations, e.g. low intelligence in migrants to a city in times of agricultural distress. Environment can shape heredity, as when persons in a more neurosis-producing environment might avoid neurosis-prone mates. It can act by *selection* as when educated families have fewer children. And it can produce differential *survival*, as in (iv).

One would expect correlations from 2 to be trivial as compared with those in 1 or 3.

3. COMMON CAUSE OPERATING ON BOTH HEREDITY AND ENVIRONMENT

(*a*) *Within Family*. Only in the exceptional instance of something productive of mutation (e.g. radiation) simultaneously affecting heredity and environment could there be anything but a general zero correlation here.

(*b*) *Between Family*. Many influences can simultaneously change environment and select for heredity between families. In a more refined analysis one could make claims for all four mechanisms noted in 1 (*a*). There can be deliberate selection by society, as when intelligent children are sent to college or delinquents to jail. Diseases may select for both environment and heredity, for instance, malaria for debility and sickle cell anaemia, and tuberculosis which selects for lower social status and leptosomatic-schizothyme constitution [Cattell 1950.] Such various influences as social placement agencies or eugenic ideals may produce similar correlations; and as Fisher pointed out [1928], dysgenic competitive-restrictive family attitudes may become correlated with higher intelligence, and other genetic qualities.

Almost every correlation observed will evidently result from several influences. Unfortunately with only four to six variables the possibility of employing factor analysis to determine the number and nature of the influences at work is remote. The allocation of contributions must therefore proceed by less systematic means. The hypothesis of the law of coercion to the bio-social mean, for example, thus proceeds from both MAVA results and findings in the field of mis-perception (projection) [Cattell, 1957].

In terms of the demand of economical experimentation for a restriction of the number of unknown *r*'s, it must be noted that elimination has to be finally carried out by taking into consideration the unknown variances needed in the solution and the combinations of equations which offer algebraic independence. What has been called the Limited Resources Design takes as variances σ^2_{we}, $\sigma^2_{we''}$ (identical twins), σ^2_{wh}, σ^2_{bh}, σ^2_{be}, and as correlations $r_{wh.we}$, $r_{we.bh}$, $r_{bh.be}$, and r''_{bhbe}. The Complete Design adds one more variance, considering $\sigma^2_{we'}$ (fraternal twins) as possibly different from the environmental variance of brothers (corrected to the same age), and it adds one correlation, $r'_{we.bh}$. This combination could probably be improved upon, for $r'_{we \cdot bh}$, as we have seen, could well be considered as r'_{wewh}. Indeed, it might be better to drop one more correlation and add one more variance namely, $\sigma^2_{bh'}$, to meet the possibility that children who are adopted are biologically selected, which we discussed at the outset. With seven unknown variances and eleven possible equations one might then give priority to four correlations, as follows: $r_{wh.we}$, $r_{bh.be}$, $r''_{bh.be}$, and $r_{bh.we}$. This assumes that $r'_{bh.we}$ is no different from $r_{wh.we}$ and $r_{wh.we'}$ from $r_{wh.we}$, but that they merely act over different hereditary and environmental ranges. It drops $r''_{wh.be}$, a placement correlation which is extremely difficult to justify, and it is only in the loss of $r_{wh.be}$ that any regret need be felt. Regarding the latter, however, it is difficult to see any possible cause of correlation other than that difference of parents, cited above (p. 197), or some oddity where extreme deviation in the family, as in Siamese twins, affects the family's fortunes while the children are still young. Nevertheless, it may be desirable to seek a twelfth independent equation to permit solution for a possibly significant value of $r_{wh.be}$.

V. The Independent Variables

The achievement of nature-nurture ratios for important personality and ability source traits is a valuable end in itself, enlightening theory and practice. But in the context of basic science it is only a means to an end—actually to two ends: an understanding of Mendelian genetic mechanisms on the one hand, and locating the chief environmental sources of behaviour modification on the other.

As to the former, we find ourselves in the area where geneticists commonly expect polygenic determination. Such writers as Wright [1960], Haldane [1924–32], Fisher [1930], Darlington [1939], Malecot [1948], Mather [1949], and Crow [1958] have developed a beautiful but still somewhat forlorn mathematical attack upon the problem of inferring the number and nature of genes from observed distributions of traits in parent and offspring generations. They have developed this biometric genetics to work upon phenotypic characters *which are only trivially modified by environment*. In the psychological field, therefore, these methods can be used only *after* the MAVA has been completed, and they will apply only to the σ^2_{wh} and σ^2_{bh} terms, the others being irrelevant.

The welding of these two methodologies into a single instrument for proceeding from psychological measurements to inferences about gene and chromosome structures, has yet to be done. The usual biometric analysis, operating on directly observed hereditary variances, as above, can compare the variance in offspring from parents of varying degrees of consanguinity and with the freedom of manipulation possible in the animal world, can make back crosses, intensive inbreeding, etc., which are impossible with human subjects. Perhaps the most that can be for seen at present for the MAVA method is the determination of σ^2_{wh} and σ^2_{bh} for (1) random and for first cousin marriages, (2) parents at a high and at a low level on a trait, and (3) populations in which social conditions make marriage more assortative or less assortative.

Yet it is hard to see any other way than the application of biometric genetics to MAVA results as a means of determining the number of genes, their dominance, epistacy, etc., in behavioural genetics. And the amount of social work involved in obtaining the cases, as well as the large numbers of cases required to estimate reliably the fine differences on which inference has to depend (See Section VIII) make this an arduous pursuit. Added to the need for further mathematical and statistical development they make this area of general methodology one of the most exacting in psychology.

In the location of the environmental sources of the environmental variance components the psychologist finds himself on more familiar and hopeful ground. It must be remembered, however, that for both nature and nurture components the MAVA method, in the context of general experimental designs, is yielding information only about dependent variables. The independent variables of heredity, the genes and chromosomes cannot be directly manipulated, at least with humans, and those of the environment cannot all be located.

It is this last point which makes the personality theories of a certain class of learning theorist so pointless. He forgets that his inability to control the 'human experiment' so as dependably to separate maturation from learning makes it impossible to test efficiently the hypotheses he so easily generates. The testing of a learning theory should go through two stages of experiment. First it is necessary to separate by the MAVA design the amount of variance due to learning from the amount due to maturation and then the former should be related to learning experiences.

But what are the independent environmental variables producing the observed dependent environmental variances? The value of σ^2_{be} is obviously caused by differences between families in social status, locality, peer group, income, sub-culture, religion, schooling, etc., and in short by whatever is shared by children in one family and distinguishes them from another. These do not lie only in the sociological setting of the family or other obvious sociological factors but may be quite subtle effects such as 'family atmospheres' generated by the interaction of personalities and partly handed down by tradition. These have been studied by Symonds [1931], Baldwin [1946] and Cattell [1950] and may be measurable by the new, objective, intra-familial attitude measures (Sealy, Sweney and Cattell, In Press). There are also such diverse contributions as the gestation environment common to the sibs, the common experience of certain family trauma and vicissitudes, etc.

What goes in σ^2_{we} will typically be all that distinguishes one child's experience within a family from another's, such as sib-position, birth trauma and illnesses, differing age and experience of parents at the critical developmental periods, and differing treatment by parents. Not all the within family environmental variance will arise from the environment within the family. Children from the same family may attend different schools, attach to different peer groups, etc.

If every independent environmental variable could be measured then the multiple correlation of these with the environmental variance contribution as measured by MAVA would reach unity. With the observed trait itself the correlation would be given in standard scores by:

$$R^2 = (\sigma^2_{we} + \sigma^2_{be}) / (\sigma^2_{bh} + \sigma^2_{wh} + \sigma^2_{we} + \sigma^2_{be}) \tag{4}$$

This equals $\sigma^2_{we} + \sigma^2_{be}$ if the test metric is first made standard scores for the general population. One possible check on the value for $\sigma^2_{we} + \sigma^2_{be}$ obtained by the MAVA method would therefore be to calculate the multiple correlation, R, between *all* environmental influences and the observed trait score which should approach $(\sqrt{(\sigma^2_{we} + \sigma^2_{be})}$ as a limit. In practice it is questionable whether a major fraction of the relevant environmental influences could be discovered and the obtained *R* value would be useful only as a lower bound for the contribution of an environmental variance.

For the population as a whole, a useful estimate of the correlation between some specific, independent variable in the environment and that part of a trait which is the purely environmental modification can be made by partialling the hereditary variance out of the observed correlation variance ratio. This assumes no correlation between hereditary and environmental influences, or treating it as that found between heredity and total environment on that trait.

Although as far as *individuals* are concerned further calculations with MAVA results are sometimes limited to coarse estimates, much precise research remains to be done using the comparisons of different *groups*. It is sometimes commented that all one has in the nature-nurture ratio is a quantitative statement which holds only for a particular culture and a particular racial mixture of population at a particular point in history. However, this very specificity provides an effective means for obtaining valuable conclusions in comparative culture and genetic studies. For example, the magnitude of environmental contribution to a trait may be related to forms of social and political organization, and genetic variance may be related to the magnitude of physically measurable racial and genotypic variance in the groups. There may be a possibility here for a whole science of comparative personality-culture relations. Of most immediate interest to the psychologist, however, is the comparison of variances and correlations for groups of different ages in the same culture, or of children in different sibship positions, etc. This is particularly discussed in section VII below. However, even the variations found by the MAVA method among $\sigma^2_{we}/\sigma^2_{be}$ ratios for factor-defined traits might solve the perennial argument between clinicians and sociologists regarding the relative importance (for a given trait and culture) of early events affecting the individual within the family and of social status and other sociological parameters broadly differentiating families.

VI. SUPPLEMENTARY CHECKS FOR MAVA RESULTS

The previous discussion of MAVA will have brought out that it has, along with other methods, a certain family resemblance to factor analysis, in that it analyses observed variances into certain abstract but theoretically important sources which cannot be directly measured. Consequently, it is not surprising that speculations have been raised as to whether factor analysis itself could assist in separating genetic and environmental contributions. The present writer [1946] hypothesised that primary personality factors could be generated either as constitutional or ' environmental mould ' traits, and awaited evidence for this. This appears to have been clearly supplied in the dynamic field by factors which on the basis of their apparent psychological character, can be classified as either drives (ergs) or sentiments (corresponding to social institution patterns). On the other hand for general personality factors this does not appear to have been demonstrated [Cattell, 1957]. It has been suggested that with some reduction of error of measurement a *U* distribution of the nature-nurture ratios might be found [Cattell, Kristy and Stice, 1959].

W. R. Thompson [1957] and the present writer [1946] have speculated that since a gene is a causal influence and usually shows, in pleiotropy, simultaneous effects on several characters, it has the properties necessary to appear as a factor. Any technically satisfactory discussion here would require more space to define factor analytic concepts such as variable density, factor-order and simple structure rotation. Here it will just be stated that in ordinary procedures it seems unlikely that a gene would be located as a factor. With the usual density of variables in psychological studies the factors would correspond to structures formed by many genes and or many environmental influences, although in exceptional cases such as sentiment structures, they might represent a single organised set of influences.

The mistake in any thinking which supposes that environment and heredity will appear, even at some higher order, respectively as just two factors, lies in the assumption that some people experience more 'total environmental' and others more 'total hereditary' contributions—for these factors by definition would have to affect all variables. One person may experience a larger contribution from a particular feature, or, generally, a subset of features, which happen to influence the psychological trait under consideration. The variance contribution from environment for a particular trait is not related to the mean impact for all features of the environment, but the weighted contribution (as in a multiple *R*), which is probably strongly influenced by some quite small subset of features different from those involved in the σ^2_{we} obtained for some other trait, or in the σ^2_{be} for the same trait.

The ordinary factor analytic model: (5)

$$P_{ji}=s_{j1}F1_i+\ \dots\ +s_{jk}F_{ki}+s_jF_{ji}$$

therefore becomes, on any likely hypothesis of hereditary and environmental contributions, one in which each *F* can be again broken down into the effect of *n* different *e*'s or environmental influences and *m* different *k*'s or genes, which appear with different weights over all personality factors, thus: (6)

$$F_{ki}=k_{e_1}e_{1i}\dots+k_{e_n}e_{ni}+k_{h_1}h_{1i}\dots+k_{h_m}h_{mi}$$

where the *k*'s are correlations (factor weights) of the given environmental or hereditary influences with the personality factor, the *e* values represent the extent of the individual's contact with that environmental influence, and the *h*'s his gene endowments (acting pleiotropically across several personality factors).

It will be recognized that this places the *e*'s and *h*'s formally in the model of second order factors, relative to the original behavioural performance measured (P_j's). Research advance along these lines may nevertheless prove impracticable because, (*a*) the requirement for factor analysis that $k>(m+1)$ may fail, since it is likely that the number of genes far exceeds the number of primary personality factors. It is possible, on the other hand, that blocks of *e*'s and *h*'s would act

VIII. SAMPLE SIZE

In an earlier exposition of the MAVA method [1960] the standard error of an abstract variance derived from observed, concrete variances, $\sigma^2{}_1$, $\sigma^2{}_2 \cdots \sigma^2{}_n$, was given by the formula

$$\hat{\sigma}^2{}_{\sigma^2} = 2k^2{}_1\sigma^4{}_1/(n_1 - 1) \ldots + 2k^2{}_n\sigma^4{}_n/(n_n - 1) \tag{7}$$

where the *k*'s are the numerical coefficients [see Cattell, 1960, p. 366] and the *n*'s the sizes of the various subsamples (twins, sibs, etc.). However, if, as subsequently proposed, the measured variance on any factor be corrected by subtracting from it the error estimated from the reliability coefficient for the given battery, namely, $\sigma^2(1-r)$, (where σ^2 is the observed test variance) then the error variance of the latter term must also be included, i.e., the terms in (7) become $2k^2{}_1\sigma^4{}_1(1-r)^2/(n-1)$ etc.

Secondly, if the standard error (or, at any rate, the confidence limits) of the ratio of variances is to be determined in order to be able to summarize the results as a nature-nurture ratio, one must take account of the presence of common terms in numerator and denominator, as follows:

$$\hat{\sigma}^2{}_R = [\hat{\sigma}^2{}_h/\sigma^2{}_e]^2[\hat{\sigma}^2{}_{\sigma^2_h}/(\hat{\sigma}^2{}_h)^2 - 2\,Cov\,\hat{\sigma}^2{}_h.\hat{\sigma}^2{}_e/\hat{\sigma}^2{}_h\hat{\sigma}^2{}_e + \hat{\sigma}^2{}_{\sigma^2_e}/(\hat{\sigma}^2{}_e)^2] \tag{8}$$

in which R is the ratio and $\sigma^2{}_e$ and $\sigma^2{}_h$ are the $\sigma^2{}_{we}$ or $\sigma^2{}_{be}$ and $\sigma^2{}_{wh}$ or $\sigma^2{}_{bh}$, or their sums, as usually derived from the concrete variances.[1] This approximation neglects higher powers of the ratio and accepts the obtained $(\sigma^2{}_{\sigma^2_e}/\sigma_{\sigma^2_h})^2$ as a substitute for the unbiased estimate which should be used. However, the magnitude of error likely to be introduced is small compared with that involved in the old formula which neglected the existence of common variance in the numerator and denominator. Practically, the result of this modification is that the confidence limits narrow to an acceptable range for a sample size of about 250 cases instead of 1000 as previously argued. Nevertheless, since empirical estimates in the ten equations solution of the within family sib variance and the between sib family variance must be experimentally independent if the above formula is to be used [Cattell, 1960] it is still necessary to measure some 2500 pairs of children. Fortunately, as eqn. (7) indicates, a shortage in one category can to some extent be made up by greater numbers in another. This has the important practical corollary that one might reduce a study to say, one hundred pairs for rarer groups such as twins or sibs reared apart and at the same time use several hundred sibs normally reared together. Although the MAVA method thus has technical promise for hitherto insoluble problems, its use will demand funds and fortitude. Yet if, as contended, behavioural science is being held up by a log jamb in behavioural genetics, and larger scale research organization is the only way to break through, the necessary social organization for research must be undertaken.

---

[1] The writer wishes to express gratitude to various colleagues, particularly H. W. Norton, for discussion and help with this problem.

PUBLICATIONS OF THE BRITISH PSYCHOLOGICAL SOCIETY

The British Journal of Statistical Psychology is issued by the British Psychological Society. The subscription price per volume is 30*s.* net for Members of the Society and 30*s.* 6*d.* (post free) for non-members. The subscription price in the U.S.A. is $4.50 (post free). Members of the Society should send their subscriptions to THE SECRETARY, British Psychological Society, Tavistock House South, Tavistock Square, London, W.C.1 : non-members to Messrs. TAYLOR & FRANCIS, LTD., 18 Red Lion Court, Fleet Street, London, E.C.4. Until further notice the Journal will be issued in two parts each year, in May and November.

Papers for publication should be sent to CYRIL BURT, 9 Elsworthy Road, Hampstead, London, N.W.3. Authors should retain copies of their manuscripts as neither the Society nor its Editors can undertake any responsibility for loss or delay in the case of unsolicited contributions. Authors will receive 25 copies of their articles free; extra copies may be purchased provided the order is given when the proofs are returned. Contributors should indicate whether they wish their reprints to be bound in covers. The printers will state current prices when sending out proofs.

Books for review should be sent to the Editor, advertisements to the Secretary, British Psychological Society.

The Society also issues quarterly *The British Journal of Psychology* and *The British Journal of Medical Psychology* : subscriptions, 60*s.* per annum for either journal, should be sent to the Cambridge University Press, Bentley House, Euston Road, London, N.W.1. The Society publishes, jointly with the Association of Teachers in Colleges and Departments of Education, *The British Journal of Educational Psychology :* for this the subscription, 30*s.* per annum, should be sent to Methuen & Co., Ltd., 36 Essex Street, London, W.C.2. Members of the Society receive the foregoing journals on special terms; enquiries should be addressed to the Secretary, British Psychological Society.

Papers for publication in *The British Journal of Psychology* should be sent to Dr. Boris Semeonoff, Department of Psychology, the University, Edinburgh; those for publication in *The British Journal of Medical Psychology* to Dr. Joseph Sandler, 96 Portland Place, London, W.1; those for publication in *The British Journal of Educational Psychology* to Professor P. E. Vernon, London University Institute of Education, Malet Street, London, W.C.1.

PREPARATION OF PAPERS

Contributors are asked to send their papers in a form which is ready for submission to the printer : that is to say, the arrangement of the material should follow that observed in previous publications of this *Journal*. Each article should be headed either with a series of numbered headings in italic, corresponding with the cross-headings of the several sections, or with a brief abstract : (the latter procedure is preferred). Each section should have a short cross-heading, in capitals; and the whole should conclude with a summary embodying the main conclusions reached. Special attention should be paid to such details as correct spelling, grammar, capitalization, numbering, etc. (particularly in the case of tables and references). In preparing manuscripts and in correcting proofs authors should conform, so far as possible, with the ' Recommendations to Authors ' set out in *The Printing of Mathematics* by T. W. Chaundy, P. R. Barrett, and Charles Batey (Oxford University Press, 1954, ch. II, pp. 21–73), or to those given in the paper on ' Setting Mathematics ' (Arthur Phillips, *Monotype Recorder*, XL, iv, 1956).

Owing to the present conditions for printing, the cost of setting up tables and algebraic formulae which require hand composition is very much greater than the cost of machine composition. Contributors are therefore advised to arrange their material so that it can be printed as economically as possible. Often numerical results and algebraic formulae can, with a little simplification, be readily adapted for purposes of machine composition. Manuscripts involving many tables and numerous equations can as a rule only be accepted if the authors are willing to pay for the additional **cost.**

In future authors will receive two complete slip-proofs of their papers, but no page proof. They will be expected to pay the cost of all corrections for which they themselves are responsible.

24 FEB 1964

BRITISH JOURNAL OF STATISTICAL PSYCHOLOGY

Vol. XVI Part 2. November 1963

CONTENTS

Page

Cronbach, L. J., Rajaratnam, Nageswari and Gleser, Goldine C. — Theory of Generalizability: A Liberalization of Reliability Theory 137

Warburton, Frank W. — Analytic Methods of Factor Rotation 165

Burt, Cyril — Is Intelligence Distributed Normally? 175

Cattell, Raymond B. — The Interaction of Hereditary and Environmental Influences 191

Notes and Correspondence — Persons Selection: A Technique for Grouping a Minimum Number of Persons to Maximally Predict a Person-Prototype 211

Book Reviews 214

Index to Volume XVI 223

Vol. XIV Part 1

May 1961

THE BRITISH JOURNAL OF STATISTICAL PSYCHOLOGY

EDITED BY
JOHN WHITFIELD

WITH THE ASSISTANCE OF
CYRIL BURT

AND THE FOLLOWING EDITORIAL BOARD

A. C. AITKEN
R. J. AUDLEY
M. S. BARTLETT
W. G. EMMETT
M. G. KENDALL
D. N. LAWLEY
E. S. PEARSON
E. A. PEEL
L. S. PENROSE
J. FRASER ROBERTS
A. RODGER
W. STEPHENSON
A. STUART
P. E. VERNON

Printed and Published by
TAYLOR & FRANCIS LTD.
RED LION COURT, FLEET STREET, LONDON, E.C.4

Price 20*s*. *per part* (*U.S.A.* $3.50) *Subscription* 30*s*. 6*d*. *per volume* (*U.S.A* $4.50)

EDNL. & PSYL. RES.

INDEX TO VOLUMES I TO XIV

An index to the first fourteen volumes of the *British Journal of Statistical Psychology*, classified according to authors and subjects, is now available. As proprietors of the journal the British Psychological Society wish to take this opportunity of expressing their gratitude to Mrs. Beatrice Warde to whose generosity and expert assistance the compilation and printing of the index is due.

OCCUPATIONAL PSYCHOLOGY

Editor: ALEC RODGER

JULY, 1963

VOLUME 37, No. 3.

CONTENTS

A Study of Frustration and Aggression Among Factory and Office Workers. By DENIS PYM
A Preliminary Enquiry into Engineering Drawing Comprehension. By J. SPENCER
Age of Manual Workers in Relation to Conditions and Demands of Work. By MICHAEL S. FEATHERSTONE and CATHERINE M. CUNNINGHAM
Traditional Values and Work Efficiency in a West African Mine Labour Force. By J. L. M. DAWSON
Prescription and Perception of the Supervisory Role. By BRIAN R. KAY
The Consistency of Meaning of Everyday Quantitative Expressions. By D. SHEPPARD
Other Books Received.

Annual Subscription **40** *shillings*

National Institute of Industrial Psychology, 14 Welbeck Street, London, W.1

INDEX TO VOLUME XVI

ARTICLES

Author	Title	Page
Brimer, M. A.	Ability and Attainment. III	112
Burt, C.	The Psychology of Value	59
Burt, C.	Ability and Attainment. II	106
Burt, C.	The use of Electronic Computers in Research	118
Burt, C.	Is Intelligence Distributed Normally?	175
Cattell, R. B.	The Interaction of Hereditary and Environmental Influences	191
Christensen, P. R. and Guilford, J. P.	An Experimental Study of Verbal Fluency Factors	1
Cronbach, L. J. Rajaratnam, N. and Gleser, G. C.	Theory of Generalizability: A Liberalization of Reliability Theory	137
Hallworth, H. J. and Waite, G.	A Factorial Study of Value Judgements among Adolescent Girls	37
Levi, S. J.	The Psychology of Value	51
Levy, P.	Ability and Attainment. IV	116
Peaker, G. F.	Abilities and Attainment. I	105
Perloff, R.	Persons Selection: A Technique for Grouping a Minimum Number of Persons to Maximally Predict a Person-Prototype	211
Warburton, F. W.	Analytic Methods of Factor Rotation	165
Whittaker, J. E.	The Psychology of Value	47
Williams, E. D.	The Psychology of Value	52
Young, M. and Gibson, J.	In Search of an Explanation of Social Mobility	27

BOOK REVIEWS

Author	Title	Page
Borko, H.	Computer Applications in the Behavioral Sciences	118
Cronbach, L. J.	Essential of Psychological Testing	222
David, F. N. and Barton, D. E.	Combinatorial Chance	126
Green, B. F.	Digital Computers in Research	214
Humphrey, G. and Argyle, M. (Eds.)	Social Psychology through Experiment	135
Lawley, D. N. and Maxwell, A. F.	Factor Analysis as a Statistical Method	219
McLeish, J.	The Science of Behaviour	129
Roberts, B. C. (Ed.)	British Journal of Industrial Relations	136
Stuart, A.	Basic Ideas of Scientific Sampling	221
Taylor, J. G.	The Behavioral Basis of Perception	127
Tomkins, S. S. and Messick, S.	Computer Simulation of Personality	221
	The Year Book of Education: The Gifted Child	217
	Information Storage and Retrieval Vol. I	222

Essentials of Psychological Testing. By L. J. CRONBACH. New York: Harper Brothers, 1961. New edition. Pp. xxii + 650. $7.00.

The first edition of Professor Cronbach's excellent textbook appeared in 1949. Since that date the development, both of techniques and of theory, has been so rapid that almost every chapter has had to be either rewritten or extensively revised. The aim and the general outline of the book, however, remain unchanged. Part I deals with 'Basic Concepts'. It begins by explaining the more specialized terms that have been found necessary in researches on mental tests, and then describes the nature and purpose of the principal types of test that are now in regular use. The section ends with practical suggestions on 'how to choose a test'.

Part II is entitled 'Tests of Ability', and starts with a long discussion of the 'Measurement of General Ability'. In common with British writers Dr. Cronbach deprecates the use of 'intelligence quotients' calculated *as* quotients, and prefers either the 'deviation I.Q.' or a 'standard scale'. After a full and impartial examination of the merits and limitations of individual tests and group tests respectively, he takes up the various problems of factor analysis. In keeping with the views so frequently expounded in this *Journal*, he regards it primarily as a technique for "classifying testable abilities". He recognizes three main classes of factor: (i) a single factor, (ii) numerous group factors of varying range, and (iii) a still more numerous array of specific factors—all classifiable in a hierarchical system. He then passes on to a detailed account of the more important kinds of 'special aptitude', and of the most efficient tests for differentiating between them.

Part III is headed 'Testing of Typical Performance'. But most of the chapters are in fact concerned with 'personality measurement' and allied topics. The volume ends with a discussion of the structure and dynamics of personality as an integral whole, emphasizing more especially the extent to which the synthetic interpretations of the experienced interviewer or clinician can elucidate, supplement, and even amend the analytic measurements furnished by quantitative tests. Dr. Cronbach concludes by assuring us that, in spite of continued criticism "psychological testing is an accomplishment of which its developers may well boast".

A bibliography, containing nearly 600 references, is appended; and each chapter includes questions and 'riders' inserted at appropriate points so that the student may check his understanding of what he has just read. The selection of tests and test material is, as he says, "confined entirely to tests and editions on the American market"; and, whereas in this country most of the work on mental testing has been, and still is, concerned with children of school age, Dr. Cronbach is more interested in the testing of adults—the area to which he himself has made notable contributions. However, so far as theory and general principles are concerned, there is no textbook on the subject which can be so confidently recommended to students both of individual psychology and of education.

CYRIL BURT.

Information Storage and Retrieval. Vol. 1, No. 1. Edited by J. FARRADANE. Oxford: Pergammon Press, 1963. Pp. 68. To private subscribers £5 per volume of four parts; U.S.A. $15.00; for libraries £10.

The growth of science has brought an enormous increase in the written records of research and theory. Even the most dedicated research worker now has difficulty in keeping abreast of current discoveries in his field of interest. Ironically enough, scientific papers on means for storing and retrieving information have been scattered among diverse publications. This new journal seeks to provide a single place for the study of these pressing problems of twentieth-century science. It is intended that papers on pattern recognition, speech analysis and relevant psychological enquiries be presented as well as those concerned with the logic and mechanics of information storage.

R. J. AUDLEY.

Some of these topics, however, are for the future. The present book provides an admirable summary of what is known and will be an indispensable starting point for further study.

M. G. Kendall.

Computer Simulation of Personality. Frontier of Psychological Theory. Edited by Silvan S. Tomkins and Samual Messick. New York and London: John Wiley & Sons. Pp. xiv + 325. 40*s*.

The general-purpose electronic computer can be used to simulate most processes, deterministic or probabilistic, provided that the elementary operations are well defined and may be expressed in symbolic form, numerical or otherwise. Hence, in so far as a theory of personality may be conceived along these lines, the use of computers in personality study is clearly of considerable value since it enables the predictions of a complex theory to be derived and evaluated.

This book does not argue the case for any one kind of personality theory, except in an introductory exercise by Tomkins, but is a collection of first attempts at simulation together with commentaries in which the limitations are usually clearly expressed. The fact that severe restrictions in this method of personality exploration do exist are, the computer enthusiast would maintain, an indication of the current vagueness and over-generality of personality theory.

However, it may be that the complexity of the subject will torment the personality simulator as much as it has the verbalizer of the problem, particularly since developing a programme of the intricacy required for a fairly complete description of personality would be an immense task on any computer. But it is hoped that something may be achieved in the description of partial processes, and one or two of the contributors to the book give indications of progress in this direction.

The papers and discussions in this book were given at a conference in which the majority of the participants were personality theorists. It is not a long volume and is divided into three parts. In the first, Tomkins puts forward a complicated scheme of interrelations within personality and Reitman examines the possibility of viewing personality as a problem solving structure. In Part 2, psychoanalytic theory is touched upon in a stimulating manner. Part 3 deals with special topics, and here Loehlin's presentation of a decision model which has feelings is especially interesting, and Uhr's discussion of models for perception and language is informative.

Although it gives no fresh insight into personality this book should appeal to all who appreciate that newer and more precise methods of exploration may solve some old problems and who wish to gain information on how these methods are used.

A. R. Pike.

Basic Ideas of Scientific Sampling. By A. Stuart. London: Charles Griffin & Co., 1962. Pp. 99. 18*s*.

Simple random sampling, stratification, clustering and multistage sampling are considered. The ideas underlying these and other sampling procedures are lucidly explained in this non-mathematical book for the non-specialist.

The exposition is novel. The author takes just one population of numbers, the heights of six individuals, and in 44 sections, each of about two pages, gives an excellent numerical treatment of sampling theory based on this particular population. At the end, the careful reader should be able

(i) to sample any population with an understanding of the principles of good design;
(ii) to calculate estimates from the sample he obtains;
(iii) to make simple inferences from these estimates.

M. Stone.

On the whole, I think this is a very useful book, and most of the suggestions I would have to make could only be satisfactorily covered by doubling its size. However, here they are:

(1) At various points the authors contrast the method of factor analysis unfavourably with the method of principal components on the grounds that, in the former, factor loadings are invariant under changes of scale, whereas in the latter they are not. This distinction, in my opinion, is largely illusory, and in practice is not a consideration which should influence the preference of a research worker for one method against the other. In factor analysis we are, in effect, seeking for a subspace and can only select factors with precision by using a supplementary convention or rotation techniques based on extraneous knowledge. In component analysis we do not need these extra assumptions, but seek for those components which explain most of the variation. The price we pay for this is that component weightings are not invariant. It is not a question of methodological superiority but of choice in assumptions.

(2) The chapter on the estimation of factor loadings refers to the fact that, with the advent of the electronic computer, the solution of the maximum likelihood equations has lost much of its terror. But not all of it. There seems to be some curious properties in the convergence of iterative solutions to the likelihood equations which have not yet been elucidated. There is also an important gap in that the book does not touch on the difficult problem of determining standard errors in factor loadings or estimates of residual variance.

An example from the book itself may make clear the kind of difficulty involved. A numerical example on p. 16 analyses observations of nine variates (which, incidentally, are not specified). A test indicates that no more than three factors are required. They contribute to the total variation 47·3 per cent, 10·7 per cent, and 4·1 per cent respectively. The residual variance is 37·9 per cent. Without any sophisticated tests, would anyone place any confidence at all in a factor contributing 4·1 per cent where the residual is nine times as large? For that matter, the second factor is not particularly strongly indicated.

(3) The chapter on the estimation of factor *scores* appears to me to evade a fundamental difficulty. The problem is that, if we wish to estimate factor loadings and factor scores, we are setting ourselves an impossible task, however large the sample, because there are more parameters under estimate than observations. The situation is essentially underdetermined. The authors (p. 88) recognize this but proceed: "We can, however, find linear functions of the x_i (the variates) which in some sense provide reasonable estimates of the f_r (the factor scores)." They do not say in what sense, but proceed to give the Thomson–Bartlett methods, which lead to the regression of the f's on the x's. Personally I am convinced that this is not a solution of the problem. Regression equations, used as estimators or predictors, will only give average results, not individual factor scores.

(4) The examples presented in the book are, naturally enough, based on psychological tests. It would have been instructive to see an example or two from non-psychological fields, for a reason which is not generally appreciated. It is customary to assume multivariate normality in test scores, at least until somebody can bring forward contrary evidence. The whole of the system of estimation developed by maximum likelihood methods depends on normality, and seeks for estimates of the parameters of underlying variables, not the observed variables themselves. It appears to be required that the selection of the sample is carried out with reference to these underlying variables, not the linear transforms of them which are actually observed, and this raises some curious problems of interpretation which may require a good deal of explanation in non-psychological applications.

Part II deals with 'The Detection of Ability and Selection for Educational Purposes'; and various writers describe the procedures used in Germany, France, Italy, Spain, Belgium, the Netherlands, Sweden, Canada, New Zealand, Africa and Japan, as well as in Britain and the United States. To readers of this journal the most interesting of the contributions in this second part will probably be those by Dr. Schaurer, Director of the *Deutsches Institut für Talentstudien* at Cologne, and by Professor Husen describing the methods used in Sweden. The latter is the only writer in this section who makes adequate use of statistical material.

Part III is concerned with the topical problem of 'Increasing the Pool of Talent'. Dr. McIntosh, Director of Education for Fife, examines the educational and social obstacles to the emergence of talent, with special references ot the United Kingdom. British readers will perhaps be surprised to discover how dissatisfied American psychologists and educationists have become over the comparative neglect of the highly talented in the present system of education: the old English scholarship is frequently cited with admiration and even at times with envy. The change we are told, is largely a consequence of the general reaction which followed the launching of the first Russian sputnik. Professor Anderson, in a candid chapter on 'the dilemmas of talent-centred educational programmes in the U.S.A.', is perhaps the most outspoken of the critics. "When it has come to seem imperative to the major emphasis on developing the nation's 'reserve of talent', as has been the case recently in the United States, then," he says, "the conflicting implications of the various loosely connected ideals become troublesome." Once again, in addition to contributions from English-speaking countries, there are notable papers by representatives from Europe, Asia and Africa, and particularly the so-called 'undeveloped countries—somewhat uneven in their merit. It is disappointing to find so little about the way the gifted pupil should be educated once he has been discovered, but perhaps that is reserved for another volume. As it is, the book forms one of the most interesting and valuable in this well-known series.

E. Moulton.

Factor Analysis as a Statistical Method. By D. N. Lawley and A. F. Maxwell. London: Butterworths, 1963. Pp. viii + 379. 21*s*.

For twenty years Dr. Lawley has been wrestling with the problems of statistical estimation and significance which arise in factor analysis; and with a considerable measure of success. This brochure, written in conjunction with Dr. Maxwell, summarizes the position as it now stands. Since much of the work which has been done to bridge the gap between the statistician and the psychologist lies in learned journals it is very useful to have a unified presentation of such authority.

Chapter 1 acts as a brief introduction and statement of the models to be discussed. Chapter 2 deals with the estimation of factor loadings by maximum likelihood. Chapter 3 covers the centroid method and Chapter 4 the method of principal components. Chapter 5 covers factor rotation and interpretation. Chapter 6 deals with the estimation of factor loadings, under various initial assumptions (for example, where a structure is laid down *a priori*). Chapter 7, a very short one, recapitulates attempts to estimate factor scores, of which more later. Finally, Chapter 8 discusses the problems of testing the difference between two populations where the existence of some factors common to both is suspected. There is an appendix on Matrices and Determinants.

The treatment, by design, has to be condensed and assumes that the reader already has a fair knowledge of theoretical statistics; for instance, that he knows the method of maximum likelihood and recognizes the Wishart distribution of variances and covariances in a multivariate normal system. Nevertheless room has been found for illustrative arithmetical examples and a few exercises. The reader will probably need to follow some of the mathematics with pencil and paper but should not encounter any real obscurities.

in most of the countries represented, and that nearly all were in general agreement with the views as now stated and amended; and, though several differed on minor points, the differences were chiefly a matter of relative emphasis. This, therefore, as Professor Wiseman has stated, would seem to be the best survey of expert opinion available on the subject at the present time.

In view of the widespread interest in higher education recently aroused by the Robbins' Report, and other discussions of kindred problems it may perhaps be helpful to summarize the conclusions here provisionally put forward. (i) Neurological, psychological and statistical evidence combines to suggest that human efficiency depends not only on training and experience but also on the individual's general and special abilities, qualities of temperament and character, and physical health, all of which rests in some measure on a genetic basis. The relative influence of the two types of factor—genetic and environmental—varies widely with variations in environmental opportunities. But in any case limitations in innate general ability (IGA or ' intelligence ') sets an upper limit to what can be achieved. (ii) Individual differences in IGA are the result partly of multifactorial and partly unifactorial inheritance. Hence, as is amply confirmed by emperical data, its distribution follows a curve of Type IV, i.e. approximately a normal curve. Assuming 15 I.Q. = 1 S.D., we obtain the following estimates for the number of children exceeding the specified borderlines: above 110 I.Q., 25 per cent; above 115, 16 per cent; above 120, 9 per cent; above 125, 4·5 per cent; above 130, 2·25 per cent; above 140, 0·55 per cent; above 150, 1·14 per cent; above 160, 0·034 per cent (more than 10 times the number we should infer from the normal curve). It is noted that the average of the children with I.Q.s over 160, one in 50 will be as far above the average of the children in the grammar school as the latter is above a mental defective. (iii) For purposes of research Terman and others defined a gifted child as the brightest in 400; later writers have preferred a lower borderline, e.g. the brightest in 10. (iv) The scores obtained in intelligence tests, especially the popular type of group test, when taken as they stand, furnish decidedly inadequate measures of IGA; in particular, tests in current use lay too much stress on ' reproductive ' and ' analytical ' thinking and too little on ' synthetic ' and ' productive or creative ' thinking. However, when checked and adjusted by supplementary data, e.g. the observations of experienced teachers who know the individual children, and further tests or interviews in the cases of discrepancy, then the results have fairly high predictive value. In the selection of gifted children, therefore, the chief value of standardized tests is to enable the judgements of individual observers (e.g. teachers) to be comparably scaled. (v) As we might expect, if mental ability is largely inherited, the *proportion* of gifted children in the ' manual classes ' is appreciably lower than in the ' professional classes ', but their *absolute number* is much greater; and of these barely one-half received the kind of education their talents deserve. (vi) On the whole, gifted children, when efficiently selected, appear to be superior to the average child in health, nervous stability and moral character, but differ widely in special abilities and in motivational characteristics. Their cultural and ethical ideals depend largely on the type of social class and home environment from which they are drawn. (vii) Judged by their performance in after-life, gifted children, selected by means of test-results (supplemented and corrected by teachers' judgements), usually do well not only during the remainder of their school careers but also in after life. When they fail, the commonest reasons seem to be lack of educational opportunities at home and at school, lack of adequate motivation, and physical or physiological changes occurring after the initial selection, particularly during adolescence.

The remainder of the *Yearbook* is divided into three sections. Part I is entitled ' Psychological Theories '. Here nearly all the different schools of thought are represented, including those of present-day Freudians and Jungians. Most of the contributors base their views principally on *a priori* arguments; but a few, notably Professors Guilford, Torrance, Pierce and Mierke, cite statistical evidence in support of their conclusions. Torrance's chapter furnishes a most useful survey of recent American work on ' creativity ', together with a bibliography of four pages.

recognition, and problem-solving and game-playing (including the description of the GPS); further possibilities for computer modelling of psychological, sociological and economic processes are mentioned.

One pressing problem at the moment is that of rapid and efficient communication between men and machines. particularly relevant in an integrated 'man-machine system' such as that for the aerial defence of the United States or that which controls manned orbital flights. A chapter is devoted to the features and problems of such systems. The author then describes his own programme 'Baseball', which is an initial attempt at providing for man-machine communication in ordinary human language.

A final two chapters introduce 'Some Computer Theory': first the logical design of computer hardware with its reminiscences of neurology; and then the logico-mathematical theory of generalized computation and computability, including a discussion of Universal Turing Machines. This last chapter is particularly notable for its clarity in what can be and often is extremely obscurely expressed.

I have described the contents of this book in detail for the benefit of those who are perhaps unaware how wide the scope of computer applications in the behavioural sciences already is. A complementary book is the recent *Computer Applications in the Behavioural Sciences* (Prentice Hall, 1962) by a number of contributors under the editorship of Dr. Harold Borko. This describes a wider range of applications than the book under review, but it is somewhat uneven; although it includes some excellent individual pieces, it does not provide such a good introduction to computer and programming fundamentals. It is to be hoped that these two books, and in particular Dr. Green's, will open many eyes to the exciting and apparently limitless applications of computers in psychology and its allied sciences. It is unfortunately the case at present, at any rate in Britain, that the computers available for research work are being unduly monopolized for the vast calculations of the nuclear physicists and their fellows, to the severe detriment of psychological applications; this is largely because of the lack of interest taken by academic psychologists—with some shining exceptions—in the subject. A year or two ago the University of London Computer Unit invited Dr. Alan Newell, one of the authors of the GPS, to lecture in London on computer simulation of human intelligent behaviour. The lectures were well attended—by programmers, mathematicians, physicists, engineers and the like; only a sprinkling of academic psychologists were present, in spite of a considerable volume of advance notification, a circumstance which can hardly be a source of pride to British psychology. One of the interesting features of 'computer psychology', by the way, is the number of its practitioners who were originally engineers, physicists, or accountants. Things are beginning to move, however; Dr. Newell's visit was not entirely without fruit, and it is to be hoped that Dr. Green's book will still further accelerate the rate of advance.

W. L. B. NIXON.

The Year Book of Education: The Gifted Child. London and New York: published in association with Teachers College, Columbia University, New York, and the Institute of Education, University of London, by Evans Brothers, 1962. Pp. xiv + 541. 63*s*.

This volume is described on the wrapper as the twenty-fourth issue of the *Year Book of Education*. It is devoted to the problem of the 'gifted child', and its purpose is to discuss what different countries are doing for those whose abilities are above the general average, and how far existing practices can be improved. There are forty contributors; thirteen are American, eighteen British, while the remainder are drawn from almost every quarter of the globe. Sir Cyril Burt was invited to write the 'General Introduction', which runs to 60 pages. This opening contribution is one of the few which relies for its conclusions chiefly on ascertained statistical data and detailed statistical analysis. We are told that the original draft was submitted to leading psychologists, geneticists and educationists, authors' intelligences in application to the problems raised by intelligence itself, and a working model, or what is the same thing, a *theory*, with descriptive and predictive capabilities, of the problem-solving behaviour of intelligent beings. Numerical calculation plays only a minor part in the operations of the GPS. An important point is that however ill-structured the phenomena, the theory itself—the computer program—is utterly precise and well-defined, as every program must be—vague generalizations and fluid definitions simply cannot be programmed, and computer programming as a methodological discipline cannot but have a good effect on a science like psychology, which has still not altogether freed itself from metaphysical and literary elements. One is at this point tempted to wonder how a member of one of the psychoanalytical schools would set about embodying his theories in computer programs.

The book under review has been written to introduce psychologists and behavioural scientists generally to the general-purpose computer and to its actual and potential applications in their own fields. The author, of the Department of Psychology at the Carnegie Institute of Technology, has the advantage of working in close touch with one of the world's leading computer centres, where indeed some of these applications have been developed with his participation. He is a computer enthusiast (computers seem to generate the same kind of loving critical devotion as railways, but without the nostalgia); indeed he warns his readers that " psychologists have a history of being fascinated by the tools of their trade, and computers are immensely fascinating . . . Of course, the only sure way to avoid the lure is to avoid programming "—as, sad to relate, too many psychologists do—" but the possibilities of a computer cannot be appreciated without first-hand experience. The best strategy is to recognize the danger and plunge ahead." His enthusiasm is safeguarded from excess by an agreeable lightness of touch and an enviable lucidity of style, even when he is in among the intricacies of actual programming. This is not a reference book: it is intended to arouse interest and overcome prejudices, in the hope that the reader will wish to ' plunge ahead ' himself.

The first five chapters, under the general heading ' Programming ', constitute one of the best general introductions to the subject for beginners from any field that the reviewer has read. They include a chapter on basic programming, that is, programming in ' machine-language ' rather than in an ' autocode '. It is nowadays fashionable to decry machine-language programming for all but professional programmers; in general, in this reviewer's opinion, this is a misplaced esotericism, and for psychologists in particular it is positively desirable that they should understand how the computer ' really works ' because of the general implications of computers for psychology already touched on. There is nevertheless, an introduction to FORTRAN programming leading up to a sample program for computing means, standard deviations and intercorrelations for data on several variables. The author also introduces the ' list-processing ' technique which he among others has helped to develop, and on which the GPS program depends. He naturally leans on his own experience with IBM computers and programming, but not so heavily that a reader unfamiliar with these would feel any discomfort. He is particularly good at conveying the 'feel' of programming, especially by describing the 'tricks of the trade' which soon become second nature to programmers with quite modest experience.

This first part describes programming particularly for numerical calculation; the next two chapters introduce the more general symbolic information-processing capabilities of computers which are as already mentioned of special relevance to psychological applications. Part III is a series of chapters devoted to applications. First, a remarkably sane and fresh discussion of ' Statistical Calculations and Data Analysis ' (see for example, the remarks on the alleged analytical objectivity of certain methods of factor rotation on p. 155). Those curious entities ' pseudo-random numbers ' are then introduced as a prelude to a chapter on the use of computers to generate schedules of experimental stimuli, and even to control their presentation in the experimental situation. A long chapter on ' Computer Models of Psychological Processes ' introduces neural net models, signal detection, pattern

BOOK REVIEWS

Digital Computers in Research. An Introduction for Behavioral and Social Scientists. By Bert F. Green. New York: McGraw-Hill, 1963. 83*s.* 6*d.*

This book, the first of its kind by a single author, is an excellent one, and deserves to be widely read both within and outside the circles to which it is addressed. The reviewer himself, a programmer in frequent touch with behavioural scientists, has read it with both pleasure and profit, and it has prompted him to set down some general reflections on the role which computers actually do and ideally should play in the behavioural sciences.

The impact of the general-purpose computer on the behavioural sciences, and on psychology in particular, though somewhat delayed, is potentially as massive and will probably be even more revolutionary than it has already been and continues to be in other fields of scientific endeavour. Perhaps because of this, the reactions of behavioural scientists to the new daemon have tended to be very mixed, ranging from outright rejection through an uneasy ambivalence between awe and suspicion to an enthusiastic and even uncritical welcome. To the astronomer, the nuclear physicist and the aeronautical engineer, the computer, though sufficiently marvellous, has come to be no more than a welcome addition to the everyday tools of his work, of which numerical calculation, the original purpose of computers and still the bread and butter of their daily work, is an essential constituent. Even in the two decades of their existence, however, it has become more and more widely realized that general-purpose computers are something more than just powerful, tireless and stupendously rapid calculating engines; there has been talk of 'mechanical intelligence', of 'electronic brains'; and psychologists have become aware, some of them with unease, that the very existence of such machines may pose for them questions they must sooner or later face.

Some psychologists have come to terms with the new machines by reducing them to their original status, that of calculating machines to which they can hand over the drudgery of their factor analyses or their analyses of variance. These do not realize, even do not wish to realize, that even as calculating machines general-purpose computers make possible not merely larger calculations of the same old kind but new types of calculation altogether. It is not nearly so widely recognized as it ought to be that the content of the familiar statistical armoury is largely determined by the limitations of manual computation; theoretical curiosities like canonical correlation analysis and canonical discriminant analysis have suddenly become practicable methods, and the peculiar characteristics of computers as calculating machines have suggested entirely new methods such as William's (*J. Ecol.*, XLVII, 1959) analysis of heterogeneity, and Sneath's (*J. Gen. Microbiol.*, XVII, 1957) taxonomical analysis. Research students in psychology and allied disciplines do not often hear of these things; indeed in some cases the computer is allowed to confirm them in their mathematical and statistical illiteracy by facilitating its application with less and less trouble to larger and larger masses of data—a sad truth familiar to all academic computer centres.

Awestruck journalistic references to 'electronic brains' which expatiate on their apparently superhuman 'intelligence' may by the natural revulsion they engender tend to obscure a kernel of real truth in the attribution of intelligence to general-purpose computers. Consider, for example, a computer engaged in translating Russian text into English; if we saw a human being doing just this we should regard him as exhibiting intelligent behaviour. Now the computer was designed by human beings and so embodies the fruits of its designers' intelligence; and in particular a human programmer prepared the Russian-English translation program which the computer is executing. Is it then so very unreasonable to deny the epithet 'intelligent' to the computer's behaviour, when we remember that the intelligence belongs to its designers and programmers? A computer executing a program is acting as an extension of the programmer's brain and intelligence in the same sort of way as microscopes or telescopes act as extensions of their users' eyes and vision. A program is a kind of projection of a segment of the programmer's intelligence which is encapsulated outside his brain and his body and which is capable of functioning quite separately from and independently of him within the computer. The computer is quite unlike the programmer's brain in certain important respects: it is not subject to any significant extent to fatigue or to error, and it can keep under simultaneous review and control very many more aspects of the problem to which it is being applied. On the other hand, computers so far do not keep within themselves any memory of the numerous jobs they are made to do, so that they present to the user a *tabula rasa* even when applied to problems they have solved on scores of previous occasions (there are, however, signs that this inability to learn from experience of their own activities may be remedied—by means of appropriate super-programs and enlarged memory capacity). A computer even now extends and augments the intelligence of the individual user; the computer of the future may be a simultaneous extension and augmentation of the intelligence of many human beings.

Programming is moreover a recursive activity—programs can be written to generate other programs. One familiar example of this recursion is the 'compiler' program which takes a programmer's program written in an 'autocode' language (e.g. FORTRAN, ALGOL, MERCURY AUTOCODE) and generates a functionally equivalent program in the basic control language of some particular machine. A more truly recursive example would be a general translation program—alas, still some way from actuality—which on being supplied with suitable data about a pair of human languages, say Russian and English, would generate a Russian-English translation program. An actual example of a second-order program of this kind is the *General Problem Solver* of Newell, Shaw and Simon ('A variety of intelligent learning in a General Problem Solver.' In *Self-organizing Systems*, New York: Pergamon Press, 1960), the main features of which are described in the book under review. One might write a computer program to prove theorems in elementary logic, another to prove the trigonometrical 'identities' with which schoolboys are plagued, and a third to solve the riddle of the missionaries, the cannibals, the river and the small boat. The point about GPS is that it can be applied to all three types of problem and many others besides; in order to cope with a specific problem of a particular type, GPS must be first supplied with the *context*—the rules of symbolic logic, for example—appropriate to the type.

GPS is remarkable enough for what it does, but its potential interest for psychologists resides even more in how this is done. Its authors put it forward not so much as a practical means for solving problems, but rather as a possible working model of *human* activity while solving them. A certain set of differential equations expressing Newton's laws of motion and gravitation is a model of planetary motions in the Solar system; a computer executing the programme for solving the equations is the model actually working. The acceptability of the astronomical model follows from the closeness with which its evolutions match those of the planets themselves; understanding of the model contributes to understanding the system modelled; confirmation of the model as a theory of planetary motion comes when it is found to predict correctly hitherto unlooked-for features of the system. Now the physical domain is matched in its order and its regularity by mathematics in the stricter sense, of which numerical calculation is the practical counterpart; but in the biological-behavioural field, except in certain specialized areas, the kaleidoscopic richness and variety of the phenomena to be reduced to order appears to defeat all attempts to emulate the successful application of mathematics in the physical sciences. The general-purpose computer is beginning to supply the deficiency. Numerical calculation is in fact a rather specialized type of machine computation, which is in general more appropriately termed *symbolic information processing*. Computer programs, the GPS for example, can be prepared as working models of complex irregularly-structured natural phenomena such as human problem-solving behaviour; the GPS is at once an encapsulated segment of its desperation that she was unable to 'make her day' because of indiscreet interruptions by, and constant clashing with, her supervisor. It will be supposed that the indiscreet interruptions and vexing clashes cannot be mitigated. In other words the situation cannot be improved by modification of the leader's behaviour. This may be attributed to her permanence in the organization for a variety of reasons; to supposed important psychological requirements of her job as buyer, to dogma based on the belief that incumbents in these jobs should have certain characteristics, or to other causes. At any rate, what concerns us here is that saleswomen who work for these persons must be able to cope with certain immutable idiosyncrasies. Now if the matrix of inter-person correlations and inter-person/person-criterion correlations is based on a variety of tests and/or items which include personality scores designed to measure the foregoing supervisory idiosyncrasies, then we might be approaching a more objective selection of saleswomen in this particular situation.

If the Wherry–Doolittle were modified as suggested, a change would have to be effected in the customary rule for deciding when one has obtained the *n*th "test". That is, in the Wherry–Doolittle, *n* is established when the $(n+1)$st "test" yields a shrunken multiple less than that acquired using *n* "tests". However, in an industrial or business organization where, for operational requirements, a fixed number of people is established for a particular section or department, the investigator would be obliged to satisfy himself with an *n* less than that which would yield the optimal multiple correlation. But this multiple should still be higher than that obtained by selecting an identical number of different persons, whose selection would be counter-indicated by the Wherry–Doolittle. If, on the other hand, operational requirements call for the selection of a number of persons greater than the number indicated by the optimal shrunken multiple, we would expect this multiple to exceed that derived from a combination of different or randomly selected people.

The identification of suppressor variables may be as important in persons-selection as in test-selection [5]. In persons selection, one might locate an individual correlating substantially with another person used as a predictor, but whose correlation with the person-criterion is negligible. This person should not be dropped from the predictor battery, thus spuriously attenuating the multiple. He might be for instance the individual who does not get along with his supervisor, but who does get along with one or more of his peers. Often when this kind of individual leaves the group there follows a reduction in the group's productivity.

It is also important to avoid confusion between person-selection with an external criterion, that is, a criterion person, and person-selection in which a criterion-person is not involved. Where we are trying to predict for a person-criterion, the same principles would prevail for person-selection as exist in test-selection. Specifically, persons would be selected who correlate highly with the person-criterion, but whose intercorrelations are relatively low. They will then not only be related to the person-criterion, but also will be differentially related to the person-criterion, each contributing his unique share to the variance of the person-criterion. Where no person-criterion is involved, selection could be made on the basis of high interpersonal correlations.

The sample size (tests) needs to be considerably larger than the number of variables (persons), otherwise the shrunken multiple would approach zero as the number of people approached the number of tests. Perhaps this problem could be handled by using carefully selected items, instead of tests, from a battery of heterogeneous tests.

Another problem concerns the nature of the criteria upon which successive samples of job applicants would be selected. This problem of generalisation to samples beyond the validation-sample, might be approached by an accumulation of data about persons which is rather broad in scope so that comparable persons, in future samples, could be objectively and easily obtained as replacements for either the group as a whole or for individual group members.

References

[1] Burt, C. (1937). Correlations between persons. *British Journal of Psychology*, XXVIII, 59–96.

[2] Lubin, A., and Summerfield, A. (1951). A square root method of selecting a minimum set of variables in multiple regression. *Psychometrika*, XVI, 271–284, 425–437.

[3] Stephenson, W. (1935). Correlating persons instead of tests. *Character and Personality*, IV, 17–24.

[4] Wherry, R. J. (1940). In Stead, W. H., and Shartle, C. P. *Occupational Counselling Techniques*. Appendix V. New York: American Book Company, pp. 245–252.

[5] Wherry, R. J. (1946). Test selection and suppressor variables. *Psychometrika*, XI, 239–247.

Hess. E. H. [1959]. The relationship between imprinting and motivation. In M. R. Jones, edit., *Nebraska symposium on motivation*. Pp. 47–77. Lincoln, University of Nebraska Press.

Higgins, J. V., Reed, W. E., and Reed, S. C. [1962]. Intelligence and family size: a paradox resolved. *Eugen. Quart.*, **9**, 84–90.

Hundleby, J., Pawlik, K. and Cattell, R. B. [In Press]. The first twenty-one general personality dimensions in objective behavioural measurements.

Hunt, J. McV. [1961]. *Intelligence and experience*. New York: Ronald Press.

Isaacs, J. T. [1962]. Frequency curves and the ability of nations. *Brit. J. stat. Psychol.*, **15**, 76–79.

Kallman, J. W. [1946]. The genetic theory of schizophrenia. *Amer. J. Psychiat.*, **103**, 309–322.

Knapp, R. B., Cattell, R. B. and Scheier, I. H. [1961]. Second order personality factor structure in the objective test realm. *J. consult. Psychol.*, **25**, 345–352.

Loehlin, J. C. [1962]. Heredity, environment and personality inventory items. Paper delivered at the APA convention in St. Louis.

Loehlin, J. C. Some corrections to Cattell's multiple abstract variance analysis method. [In Press.]

Malecot, G. [1948]. *Les Mathematiques de l'heredité*. Paris, Masson.

Mather, K. [1949]. *Biometrical genetics*. New York: Dover Publ.

Penrose, L. S. [1959]. *Outline of Human Genetics*. New York: Wiley.

Piaget, J. and Inhelder, B. [1956]. *The child's conception of space*. London: Routledge and Kegan Paul.

Sealy, A. P. and Cattell, R. B. The objective measurement of intrafamilial attitudes. [In Press].

Spillane, J. and Phillips, R. [1937]. Huntington's chorea in South Wales. *Quart. J. Med.*, 6403–423,

Symonds, P. M. [1931]. *Diagnosing personality and conduct*. New York: Appleton Century.

Sweney, A. B. and Cattell, R. B. [1962]. Relationships between integrated and unintegrated motivation structure examined by objective tests. *J. Soc. Psych.*, **57**, 217–226.

Thompson, W. R. [1957]. Traits, factors and genes. *Eugen. Quart.*, **4**, 8–16.

Vandenberg, S. [1961]. The hereditary abilities study: hereditary components in a psychological test battery. *Amer. J. human Genet.*, **14**, 220–237.

Wright, S. [1960]. Path coefficients and path regression: alternative or complimentary concepts? *Biometrics*, **16**, 189–202.

Vol. XVI
Part 2

The British Journal of Statistical Psychology

November
1963

NOTES AND CORRESPONDENCE

PERSONS SELECTION: A TECHNIQUE FOR GROUPING A MINIMUM NUMBER OF PERSONS TO MAXIMALLY PREDICT A PERSON-PROTOTYPE

By Robert Perloff
Purdue University

This note introduces a statistical technique that may be of use in solving problems of human relations. The concepts and procedures of test selection methods such as the Wherry–Doolittle [2, 4] are combined with the notion of correlations between persons, developed by Burt [1] and Stephenson [3]. The combination aims at selecting the smallest number of persons to yield a maximal multiple correlation with an external criterion, where the criterion, instead of being a test, a rating, or any other measure of job success, is a person. This person may be viewed generically as a leader, and specifically, according to local circumstances, as a supervisor, foreman, or committee chairman.

Harmony between supervisor and worker or worker and worker is very important when comparatively small groups of persons are required to act together as a team. Furthermore, the selection of individuals for the purpose of forming a good team may become increasingly important as automation reduces the number of workers involved in any task. Under these conditions attention ought to be given to the selection of a number of individuals from a larger applicant battery. The selected individuals should (*a*) reflect or correlate with a person-criterion or composite-person-criteria for appropriate personality, attitude, achievement, or biographical characteristics, and (*b*) should reflect one another, without regard to their composite relation to a person-criterion.

In the case where the criterion takes the form of a person rather than a test or trait, the appropriate modifications of the Wherry–Doolittle procedure should provide a quantitative yardstick which the investigator may use to make decisions regarding an efficient grouping of individuals. An illustration might be a situation in which management is establishing a new working unit which will comprise a supervisor and perhaps a handful of subordinates. Let it be supposed that its members work closely with one another and with the unit leader, and that incompatibility of the subordinates with the leader would vitiate the work of the unit and reduce the effectiveness of ordinary selection techniques. On the other hand, where a group of workers is not in frequent contact with a supervisor, or where this contact is regarded as not being crucial to the productivity of the group, the members might be selected in terms of their homogeneity with respect to appropriate characteristics, in addition to their performance on standard tests or their measures on certain biographical indices. Thus conventional selection instruments can be supplemented by various interpersonal profiles derived from a battery of job applicants.

The central problem is the selection of a minimum set of variables (in this case persons rather than tests) in what is essentially a multiple regression problem for the best ‘prediction’ of an ideal person taken as the criterion. There is a need for some such device in several types of human relations research: aside from any extrinsic or methodological considerations, one might consider how the techniques could be used as an aid in the solution of several kinds of human relations problems.

Consider the friction and resulting grievances that are commonplace among small groups of employees. For example, it is not unusual to hear a saleswoman exclaim in

References

Anastasi, A. [1959]. Differentiating effect of intelligence and social status. *Eugen. Quart.*, **6,** 84–91.

Baldwin, A. H. [1946]. Differences in parental behaviour toward three and nine year old children. *J. Person. Res.*, **15,** 143–165.

Burt, C. [1912]. The inheritance of mental characteristics. *Eugen. Rev.*, **4,** 1–33.

Burt, C. [1946]. Eye-colour and defective colour vision. *Eugen. Rev.*, **37,** 149–156.

Burt, C. [1952]. *Intelligence and Fertility.* London: Cassell.

Burt, C. and Howard, M. [1956]. The multifactorial theory of inheritance. *Brit. J. stat. Psychol.*, **9,** 115–125.

Burt, C. and Howard, M. [1957]. The relative influence of heredity and environment on assessments of intelligence. *Brit. J. stat. Psychol.*, **10,** 99–104.

Burt, C. [1961]. Intelligence and social mobility. *Brit. J. stat. Psychol.*, XIV, 3–25.

Cattell, R. B. [1946]. *The description and measurement of personality.* New York: World Book Co.

Cattell, R. B. [1950]. *Personality: a systematic, theoretical and factorial study.* New York : McGraw Hill.

Cattell, R. B. [1953]. Research designs in psychological genetics with special reference to the multiple variance analysis method. *Amer. J. human Genet.*, **5,** 76–93.

Cattell, R. B. [1957]. *Personality and motivation structure and measurement.* New York: Harcourt, Brace and World.

Cattell, R. B. [1960]. The multiple abstract variance analysis equations and solutions: for nature-nurture research on continuous variables. *Psychol. Rev.*, **67,** 353–372.

Cattell, R. B. [1961]. Group theory, personality and role. Chap in *Defence Psychology.* (Edit. T. A. Geldard.) Oxford: Pergamon Press.

Cattell, R. B. [1963]. The theory of fluid and crystallized general intelligence: a crucial experiment. *J. educ. Psychol.*, **54,** 1–22.

Cattell, R. B., Blewett, D. B. and Beloff, J. R. [1955]. The inheritance of personality: a multiple variance analysis of approximate nature-nurture ratios for primary personality factors in Q-data. *Amer. J. human Genet.*, **7,** 122–146.

Cattell, R. B., Breul, H. and Hartman, H. P. [1952]. An attempt to more refined definition of the cultural dimensions of syntality in modern nations. *Amer. Soc. Rev.*, **17,** 408–421.

Cattell, R. B., Kristy, N. and Stice, G. F. [1957]. A first approximation to nature-nurture ratios for eleven primary personality factors in objective tests. *J. Abnorm. Soc. Psychol.*, **54,** 143–159.

Cattell, R. B. and Scheier, I. H. [1961]. *The meaning and measurement of neuroticism and anxiety.* New York: Ronald Press.

Crow, J. T. [1958]. Some possibilities for measuring selection intensity in man. *Hum. Biol.*, **30,** 1–13.

Darlington, C. D. [1939]. *The evolution of genetic systems.* Cambridge: Cambridge University Press.

Eells, K., Davis, A. et al. [1951]. *Intelligence and cultural differences.* Chicago: University of Chicago Press.

Eysenck, H. J. [1953]. *The structure of human personality.* London: Methuen.

Fuller, J. T. and Thompson, W. R. [1930]. *Behaviour genetics.* New York: Wiley.

Fisher, R. A. [1930]. *The genetical theory of natural selection.* London: Oxford University Press.

Gorsuch, R. L. and Cattell, R. B. The second order structure of personality on the questionnaire medium. (In press).

Haldane, J. [1924–32]. A mathematical theory of natural and artificial selection. *Proc. Camb. philos. Soc.*, **23**: 19, 158, 363, 838; **26**: 220; **27**: 131, **28**: 244.

THE BRITISH JOURNAL OF STATISTICAL PSYCHOLOGY

32

EDITED BY
JOHN WHITFIELD

WITH THE ASSISTANCE OF
CYRIL BURT

AND THE FOLLOWING EDITORIAL BOARD

A. C. AITKEN
R. J. AUDLEY
M. S. BARTLETT
W. G. EMMETT
M. G. KENDALL
D. N. LAWLEY
E. S. PEARSON
E. A. PEEL
L. S. PENROSE
J. FRASER ROBERTS
A. RODGER
W. STEPHENSON
A. STUART
P. E. VERNON

PSYCL. RESEARCH
LIBRARY

VOLUME XIV

Printed and Published by
TAYLOR & FRANCIS LTD.
RED LION COURT, FLEET STREET, LONDON, E.C.4

Vol. XIV
Part 1
The British Journal of Statistical Psychology
May
1961

A NOTE ON GUTTMAN'S LOWER BOUND FOR THE NUMBER OF COMMON FACTORS[1]

By HENRY F. KAISER
University of Illinois

Guttman's classic lower bound for the number of common factors is extended to the completely general case where communalities may lie in the closed interval from zero to one.

Guttman [2] has established a lower bound for the number of common factors. His result is as follows. Given R, the correlation matrix of order n, let $R_0 = R - I$, and G^2 = the diagonal matrix of squared multiple correlations of each variable with the remaining $n-1$: then the number of non-negative latent roots of $R_0 + G^2$ is a lower bound to r, the rank of the matrix $R_0 + H^2$, where H^2 is a diagonal matrix of communalities. Guttman's treatment is for the usual case : $0 < h^2_j < 1$, $j = 1, 2, \ldots, n$. It is the purpose of this note to extend his result to the general case, viz. $0 \leqq h^2_j \leqq 1$, $j = 1, 2, \ldots, n$.

Let the *positive index* of a real symmetric matrix be the number of positive latent roots of the matrix, and similarly, the *non-negative index* the number of non-negative latent roots. The crux of Guttman's finding lies in the following lemma.

Lemma. Let B be any real symmetric matrix, C a non-singular real Gramian matrix, and A their sum, so that

$$A = B + C \; ; \tag{1}$$

and let a be the positive index of A, and b the non-negative index of B : then

$$a \geqslant b. \tag{2}$$

We write the identity

$$(R_0 + H^2) = (R_0 + G^2) + (H^2 - G^2). \tag{3}$$

Since $g^2_j < h^2_j$ for $0 < h^2_j < 1$, $j = 1, 2, \ldots, n$ ([1], pp. 92–93), then, letting k^* be the non-negative index of $(R_0 + G^2)$, we may associate $(R_0 + H^2)$ with A, $(R_0 + G^2)$ with B, and $(H^2 - G^2)$ with C in Guttman's Lemma, thus

$$r \geqslant k^*, \tag{4}$$

Guttman's lower bound for r.

In general, however, we may only say $g^2_j \leqslant h^2_j$ [3], and the Lemma is then inapplicable since $(H^2 - G^2)$ is no longer necessarily non-singular. We shall now show that Guttman's conclusion holds for the general case, except when $R_0 + G^2$ is singular.

---

[1] This research was supported by the United States Air Force under Contract No. AF41 (657)–76, Professor Charles Wrigley, principal investigator.

Let $D = dI$ be an arbitrary positive scalar matrix of order n. We rewrite (3) as

$$(R_0 + H^2) = (R_0 + G^2 - D) + (H^2 - G^2 + \mathrm{D}), \tag{5}$$

and since $(H^2 - G^2 + D)$ is necessarily positive definite, we may apply the above Lemma, and say

$$r \geqslant k_d^* \tag{6}$$

where k_d^* is the non-negative index of $(R_0 + G^2 - D)$. If m_s is any latent root of $(R_0 + G^2)$, then $m_s - d$ is the corresponding latent root of $(R_0 + G^2 - D)$. By choosing d less than the smallest positive latent root of $(R_0 + G^2)$, then

$$k = k_d \tag{7}$$

where k and k_d are the positive indices of $(R_0 + G^2)$ and $(R_0 + G^2 - D)$ respectively. (Unfortunately, we cannot in general choose d such that $k^* = k_d^*$ since $(R_0 + G^2)$, if singular, would have a latent root or roots equal to zero.) Obviously $k_d^* \geqslant k_d$, and using (6) and (7),

$$r \geqslant k.$$

Thus, for all appropriate H^2, the number of positive latent roots of $(R_0 + G^2)$ is a lower bound for the number of common factors.

Guttman [2], in his original paper, has applied his Lemma to show that if we have $\hat{g}^2{}_j < g^2{}_j$, $j = 1, 2, \ldots, n$, weaker lower bounds to the $h^2{}_j$, then $\hat{k}^* \leqslant k^*$, a weaker lower bound to r for $0 < h^2{}_j < 1$. Through a development similar to that of the preceding paragraph, we may generalize his result so that if $\hat{g}^2{}_j \leqslant g^2{}_j$, $j = 1, 2, \ldots, n$, then $\hat{k} \leqslant k$, a weaker lower bound to r for all H^2.

REFERENCES

[1] GUTTMAN, L. (1940). Multiple rectilinear prediction and the resolution into components. *Psychometrika*, V, 75–99.

[2] GUTTMAN, L. (1954). Some necessary conditions for common-factor analysis. *Psychometrika*, XIX, 149–161.

[3] GUTTMAN, L. (1956). 'Best possible' systematic estimates of communalities. *Psychometrika*, XXI, 273–285.

Vol. XIV
Part 1

The British Journal of Statistical Psychology

May
1961

INTELLIGENCE AND SOCIAL MOBILITY

By CYRIL BURT
University College, London

The main thesis of the following paper is that, in a highly organized society, the discrepancies between the general intelligence of the children and the occupational class into which they are born is bound to produce a large and fairly constant amount of ' basic mobility ', quite apart from any deliberate changes in the political or educational structure of the society.

Since the correlation between the intelligence of fathers and sons is only about 0·50, it is evident that, when classified according to their occupational status, (i) the mean intelligence of the children belonging to each class will exhibit a marked regression towards the general mean, and (ii) the intelligence of the individual children within each class will vary over a far wider range than that of their fathers. These deductions are fully confirmed by tables compiled to show the actual distribution of intelligence among adults and children belonging to the various occupational categories. It follows that, if the frequency distribution within the several classes is to remain constant (and still more if there is to be an increasing degree of vocational adjustment among later generations), a considerable amount of social mobility must inevitably take place, involving between 20 and 30 per cent of the population. Approximate estimates are attempted of both the actual and the ideal amounts. Data obtained from the after-histories of schoolchildren, followed up in later life, are analysed to ascertain the main psychological causes tending to produce a rise or drop in occupational status.

I. THE CLASS DISTRIBUTION OF INTELLIGENCE

Aim. In the course of a recent discussion on the mental differences between social classes [19, 20, 21] I argued that the apparent differences between the class-means for general intelligence were to be explained partly by the effects of social mobility in transferring abler individuals from lower classes to higher and duller individuals from higher classes to lower, and partly by the manner in which inherited or innate differences are transmitted from one generation to another. Several sociological writers, however, have questioned both these suggestions, or at any rate the way in which I assumed the two processes had actually operated. In this paper, therefore, I propose to offer more detailed evidence to support the interpretations I put forward, and at the same time to answer, so far as I can, the various objections raised against the arguments which I advanced on these various points ([19], pp. 22f, section on ' Social Mobility ').

The data which I shall analyse are drawn from two overlapping inquiries, or rather two series of inquiries: (i) cross-sectional surveys of pupils in London schools, initiated primarily for the purposes of educational or vocational guidance and selection; (ii) longitudinal studies of backward, gifted, and normal pupils, followed up into adult life chiefly to check the accuracy of the assessments and recommendations made while the children were still at school [4, 26]. The surveys and the subsequent inquiries were carried out at intervals over a period of nearly fifty years, namely, from 1913 onwards; and much of the data is due to the willing cooperation of numerous collaborators, particularly teachers and social workers in the service of the London County Council and colleagues or senior research-students working for the National Institute of Industrial Psychology, to all of whom I am deeply indebted.

Points of Agreement. It may help to clarify the issues involved if I begin by summarizing the major points on which both my critics and myself would, I fancy, be in general agreement.

1. During the period covered by our inquiries the population, from which our samples are drawn, and to which we intend our conclusions to apply, greatly increased in numbers, though at a diminishing rate. Both the increase itself and the diminution in the rate of increase were a continuation of processes that had been going on during the preceding half-century. Thus in 1860 the total population of England and Wales was nearly 20 million; in 1910 it was almost twice as large—36 million; and in 1960 it was 45 million [10, 11, 24].

2. During the last half-century the proportional number of children in the population steadily declined and that of the elderly steadily increased. In 1910 31 per cent of the population were boys and girls of school age (i.e. under 15) and 7 per cent men or women over 60; in 1960 only 22 per cent were of school age and 14 per cent over 60.

3. Among the lower working classes (unskilled manual labourers) both the birthrate and the deathrate were appreciably higher than among the semi-skilled, skilled, or professional classes. The number of live births per married woman averaged about 3·8 among unskilled labourers and only 1·8 in the professional classes. The differences in the mortality rates were much smaller, averaging 13·1 per 1000 among the unskilled labourers and only 10·8 per 1000 in the professional classes. For birth and death alike the absolute rates and the class differences have both appreciably diminished during the period in question. The excess of birthrate over the deathrate has been by far the most important cause of the increase in the population [11, 24].

4. During the period for which information is available there has been no great change in the average level of general intelligence. The results of the second Scottish survey indicated an actual improvement in the average score with the tests employed [7]; but I myself believe, as Sir Godfrey Thomson suggested in his preface to the report, that this was an artificial and somewhat misleading result, due partly to increased familiarity with the tests and methods of testing [8]. On the whole, a survey of the relevant evidence would appear to suggest an actual but comparatively slight decline during the period in question, approximating to a drop of 1 or 2 I.Q. points per generation [5].

5. The amount of individual variation about the average level of intelligence has apparently remained fairly constant; certainly it has not declined [6, 16]. When we compare the printed tables giving the standard deviations for complete age-groups, we usually find that in the earlier surveys it is about 12 or 13 I.Q. points and in the later as much as 16 or even more. But the apparent increase is in all probability to be explained by the fact that the later test-scales are greatly improved, and as a result decidedly more discriminative. Where the same tests have been used the figures show no appreciable change.

6. There are appreciable differences in the average level of intelligence in the different socio-economic classes, and in spite of the remarkable improvements in material and cultural conditions, the differences have altered hardly at all during the period in question [1, 3, 16].

On all these points further research and more exact information is undoubtedly required. But I believe that, as a rough provisional statement, what I have said would be accepted by most social psychologists.

Points of Disagreement. The issues on which disagreement has been most strongly expressed are those relating to the genetic hypothesis. Dr. Floud and Dr. Halsey, for example, deny that the apparent differences between the class-means for general intelligence are in any degree due to innate differences; and both contend instead for " a hypothesis of near-randomness in the social distribution of innate intelligence ". This implies that the means for all the classes would be approximately the same. Many of their colleagues have also argued that even " the apparent differences in intelligence between individuals, whether adults or children, result not from genetic causes but solely or mainly from environmental conditions ". Dr. Halsey, however, is prepared to admit that individuals may vary in innate ability; but the model he has put forward to explain how such differences are in his view transmitted and redistributed diverges widely from mine [21]. In particular he criticizes both the amount of social mobility which I had assumed and the length of time over which I assumed it had operated; his view, like that of many other social writers[1], apparently supposes that social mobility is a comparatively late phenomenon, the result more especially of recent social and educational reforms.

II. ALTERNATIVE METHODS OF ANALYSIS

Correlational Analysis and Variance Analysis. Many of the foregoing criticisms arise, I fancy, very largely from the fact that the method which I adopted in the investigations cited differed considerably from those adopted for psychological researches on heredity in the past. Most psychologists have discussed the problems of genetics in terms of the correlational procedures popularized by Karl Pearson; the investigations of my coworkers and myself were based mainly on an analysis of variance using the techniques applied by Ronald

---

[1] Most of them ignore genetic influences altogether: see the interesting papers in *Population Studies*, IX, pp. 72–81, 82–95, XI, pp. 123–136, 262–8.

Fisher. Unfortunately such methods still seem unfamiliar to the majority of psychologists and sociologists working in this country.

The analysis of variance has numerous merits with which statistical investigators are already well acquainted; but in the field of genetics it has one special advantage over the older correlational techniques. In nearly all the earlier statistical studies heredity itself was commonly conceived as 'the tendency of like to beget like'; hence the correlation coefficient, as a measure of likeness, seemed the obvious tool. On the Mendelian theory, however, genetic influences are responsible, not only for resemblances between members of the same family, but also for differences, i.e. for individual variations. Genetic variability within families receives little or no attention from those psychological critics who still accept the Pearsonian view, and think mainly in terms of correlations; yet, as we shall see in a moment, variability within families forms one of the chief causes of mobility. Moreover, it was to a large extent the exclusive reliance on correlational analysis which was responsible for the abnormally low estimates which Pearson and his followers reached for the influence of environment. Variance techniques make it far easier to give due weight to environmental influences, and to the further complications which result from the fact that environment and heredity so often work in the same direction.

In the present paper I shall, so far as possible, avoid unfamiliar methods and formulae. Nevertheless, because I believe that researches undertaken in the near future should be deliberately planned to permit the application of these newer and more efficient techniques, I will first attempt a brief explanation of the type of procedure that would seem most appropriate, and at the same time indicate how it is related to the more familiar correlational procedures. This may to some extent help to elucidate several of the points in my recent paper which the critics have either questioned or misunderstood.

The Factorial Analysis of Variance. Let us start, rather on the lines of Fisher (*op. cit. inf.*, pp. 210f.), with the simplest type of situation—that in which only two independent components of variance are involved. To make the problem concrete let us consider the case of n identical twins, reared in different environments, and tested or assessed for some form of educational or occupational efficiency, which we may plausibly suppose to be the result of both genetic (g) and environmental factors (e). If x_i denotes the assessment for the ith individual, r a correlation coefficient, and s^2 an estimated variance, we may write

$$x_i = g_i + e_i \qquad (i = 1, 2 \ldots n) \qquad \text{(i)}$$

Squaring and summing we obtain

$$\Sigma x^2{}_i = \Sigma g^2{}_i + 2\Sigma g_i e_i + \Sigma e_1{}^2,$$

and therefore $s^2{}_x = s^2{}_g + 2r_{ge} s_g s_e + s^2{}_e,$

$$\text{or } s^2{}_x = s^2{}_g + s^2{}_e, \qquad \text{(ii)}$$

if it can be assumed that the foster-homes have been chosen in a way quite unrelated to the intellectual level of each child's own family.

If we adopt a correlational procedure, the appropriate coefficient will be the *intra-class* correlation. Accordingly, following Fisher (with a slight change of notation) let us put

$$r_{ii} = \frac{G}{G+E}, \tag{iii}$$

where G and E denote the genetic and environmental variances respectively. Now, as Fisher shows, we can obtain an unbiased estimate of r_{ii} from the two equations

$$n(1-r)s^2_x = \sum(x_i - \bar{x}_j)^2 = ns^2_w, \qquad (i=1, 2 \ldots 2n) \tag{iv}$$

$$(n-1)\,(1+r)s^2_x = 2\sum(\bar{x}_j - \bar{x})^2 = (n-1)s^2_b, \qquad (j=1, 2 \ldots n) \tag{v}$$

where x_i (as before) denotes the assessment of the *i*th individual, $\bar{x}_j$ the mean of the family to which that individual belongs, s^2_x the estimated total variance, s^2_w the variance ‘ within families ’, and s^2_b the variance ‘ between families ’, n the number of families, and $2n$ therefore the number of twins, and r the correlation between the twins’ intelligence[1].

To estimate the relative size of the contributions of G and E, however, we need two further equations. Fisher shows in the course of his discussion that

$$\sum^{mn}(x - \bar{x}_j)^2 = n(m-1)E, \tag{vi}$$

$$\text{and } \sum^{n}(\bar{x}_j - \bar{x})^2/(n-1) = G + E/m, \tag{vii}$$

where m is the number in each family (with twins $m=2$). Substituting from (iv) and (v), and then solving for $G = s^2_g$ and $E = s^2_e$, we obtain

$$G = \tfrac{1}{2}(s^2_b - s^2_w), \tag{viii}$$

$$E = s^2_b. \tag{ix}$$

These then are the equations we require; and I have ventured to call the whole procedure the ‘ factorial analysis of variance ’.[2] On substituting in equation (iii) from equations (viii) and (ix) we have for the coefficient of correlation

$$r_{ii} = \frac{s^2_b - s^2_w}{s^2_b + s^2_w}, \tag{x}$$

which sums up the relation between the results of the two alternative procedures —correlation and analysis of variance—in the simplest conceivable case.

In the foregoing problem—that of identical twins brought up in separate homes—both the variance within the family due to genetic influences and the effects of environment so far as it operates in the same direction as the genetic influences could be safely ignored. If we wish to estimate the former, we can take ordinary siblings brought up from birth in the *same* environment, e.g. orphanages and other residential institutions (cf. [3], pp. 90–91); and if we want to determine the effects of any correlation between environment and heredity, we can either reintroduce the correlational term r_{ge} or calculate the additional variance directly by the method described in the earlier paper [14].

---

[1] Fisher, *Statistical Methods for Research Workers* (5th ed., 1934, chapter VII, ‘ Intra-Class Correlation and Analysis of Variance ’. The formula for r is given on p. 212, and the equations for the sum of squares ‘ within families ’ and ‘ between families ’ will be those given in Table 39 not Table 38: cf. also [4], pp. 675f. and [14], pp. 106f.

[2] This type of analysis has wide applications in psychometrics, and may also be extended to the study of interactions: see, for example, this *Journal*, VIII, p. 116.

Multiple Cross-Classifications. To deal with more complex situations the foregoing techniques can readily be extended to problems involving a multidimensional classification. In such cases, it may be noted, both the algebraic solutions and the arithmetical calculations become much simpler if the successive classifications are dichotomous. Thus, in dealing with genetical problems, it would certainly be desirable to take into account temperamental and motivational tendencies (m) as well as cognitive abilities (a). Since a general factor underlies each, we may for most purposes, treat each as supplying the basis for a further two-fold classification. In studying the influence of social class we must cross-classify both the genetic factors and the environmental according to the variations in family (f) and in social class (c): this would mean that our simple dichotomous equation (equation (i) above) must now be rewritten

$$x = g_c + g_f + e_c + e_f, \tag{xi}$$

and, if we wish to include motivational factors as well as cognitive, we must be prepared to work with eight variables, g_{ac}, g_{mc}, g_{af}, etc. In either case the derivation of the formulae will proceed much as before. But, by deciding in advance which variables we will include and which we will exclude (e.g. by arranging to keep certain conditions constant) and by avoiding so far as possible interactions or intercorrelations between the variables retained, many of the complications may be eliminated or at least reduced to insignificance.

Incidentally, we may note that with an analysis of variance it is not necessary (as it would be with a correlational procedure) to assume that the social classes themselves must be expressed by measurable quantities (e.g. by income) or ranked in linear order (e.g. in terms of prestige), as in fact has been the custom in many sociological inquiries. The method thus avoids the difficulty experienced by investigators of mobility who have endeavoured to assess or rank rural occupations on the same scale as urban and industrial[1]. The calculations would become simpler still if we were content, with several recent investigators (Lipset and Bendix [23], for example), to reduce the occupational classification in either case to a twofold division, namely, manual and non-manual.

---

[1] Glass [13], for example, in his study of social mobility in Britain, attempts to classify urban and rural populations together, and in this he is followed by L. Livi and K. Svalastoga in Italy and Denmark respectively—agricultural countries where a unidimensional classification of this kind ceases to be plausible [9]. Most other investigators classify the urban and the rural populations separately [15, 17, 18]. Even in the case of urban occupations much of the data available is expressed in terms of occupational categories which it would be very difficult to rank; in such cases, therefore, the analysis of variance, or (if Pearsonian techniques are preferred) the calculation of contingency coefficients, is far more appropriate than the calculation of correlations.

In principle, therefore, the questions with which we are concerned are essentially problems in multivariate analysis. Multivariate analysis can take several forms; and, according to the specific nature of the question we wish to answer, we may use either factor analysis, regression analysis, or discriminant analysis. But, once again, in any future research it is desirable that the investigator should keep explicitly in mind from the outset the kind of statistical techniques that are suitable for his problem and his data, and having made his choice, plan his inquiry accordingly.

In the following discussion, which is intended merely as a pilot inquiry, I shall, to begin with, confine myself primarily to assessments for general intelligence and leave motivational factors to a later section. I shall compare assessments for adults and children drawn always from the same families; but I shall adopt a moderately elaborate occupational classification. The data are too crude and limited for a detailed examination by a full analysis of variance. Moreover, in this paper it is my purpose to keep, so far as possible, to the simplest and most intelligible methods of comparison, relying largely on the percentage methods favoured by sociologists themselves. But the differences revealed, I fancy, will be sufficiently striking to lend strong support to the conclusions drawn.

III. Frequency Distributions for Adults and Children

Sources of Data. In studying the distribution of intelligence among the different occupational classes it is in my view desirable to examine, not only (as is usually done) the class-means, but the entire frequency distributions. Accordingly in Tables I and II I give frequencies both for adults and for children. For the children the bulk of the data was obtained from the surveys carried out from time to time in a London borough selected as typical of the whole county. The methods by which the assessments for intelligence were made have been described in earlier papers and in L.C.C. Reports [3, 5, 16]. For the boys who belong to the highest occupational classes, drawn for example from families who would not ordinarily send their children to Council schools, much of the data was collected in the course of work on vocational guidance at the National Institute of Industrial Psychology. The data for the adults was obtained from the parents of the children themselves. Usually our more immediate purpose was to secure practical estimates of both the average level and the range of intelligence required in the commoner types of occupation. In addition, however, when working with backward children we often wanted to see how far the backwardness was a family characteristic. And at all levels an incidental aim was to secure material for studying the problem of mental inheritance. For obvious reasons the assessments of adult intelligence were less thorough and less reliable.

The Occupational Classification. The occupational classification is much the same as that used in previous reports. It has been described by Carr-Saunders and Caradog Jones in their book on *Social Structure in England and*

Wales ([2], Table XXXI, p. 56). Unlike the classification used in the more recent studies of social mobility it is based, not on prestige or income, but rather on the degree of ability required for the work. Class I includes those engaged in the highest type of professional and administrative work (university teachers, those of similar standing in law, medicine, education, or the church, and the top people in commerce, industry, or the civil service); class II consists of those engaged in lower professional or technical work (including most teachers, men of business, and executive clerks in the higher grades); class III of those working in intermediate types of clerical, commercial, or technical work; class IV includes those ordinarily classified as skilled workers, but it also contains an appreciable number who are engaged in commercial or industrial work of an equivalent level; class V consists of semi-skilled workers and those holding the poorest type of commercial position; class VI of unskilled labourers, casual labourers, and those employed on coarse manual work. It will be noted that the numbers in the higher groups or classes are far smaller than those in the lower. These subdivisions were in fact chosen because at the outset of our work we had in mind the proportionate numbers of children (*a*) who were transferred to Central Schools (about 12 per cent), (*b*) who were awarded junior county scholarships and transferred to what were then called secondary (i.e. grammar) schools (about 3 per cent), and (*c*) who were of exceptionally high intelligence and for the most part in attendance, not at a council school, but at one of the older public schools or at a preparatory school of similar type (about 0·3 per cent); and we wanted the occupational classification to tally so far as possible with the educational classification.

In constructing the tables the frequencies inserted in the various rows and columns were proportional frequencies and in no way represent the number actually examined: from class I the number actually examined was nearer a hundred and twenty than three. To obtain the figures to be inserted (numbers per mille) we weighted the actual numbers so that the proportions in each class should be equal to the estimated proportions for the total population. Finally, for purposes of the present analysis we have rescaled our assessments of intelligence so that the mean of the whole group is 100 and the standard deviation 15. This is done because the results of so many intelligence tests nowadays are expressed in terms of conventional I.Q.'s conforming to these requirements.

IV. AMOUNT OF MOBILITY

The Distribution of Adults. From the figures set out in the last column of Table I it will be seen that there are appreciable differences between the average levels of intelligence in the various classes. The average for the highest class of all—those holding the highest professional or administrative appointments—is practically 40 I.Q. points above the general level. The differences between the means for the last three classes are much smaller, largely because the numbers are far greater.

TABLE I. DISTRIBUTION OF INTELLIGENCE ACCORDING TO OCCUPATIONAL CLASS: ADULTS

	50– 60	60– 70	70– 80	80– 90	90– 100	100– 110	110– 120	120– 130	130– 140	140+	Total	Mean I.Q.
I. Higher Professional									2	1	3	139·7
II. Lower Professional							2	13	15	1	31	130·6
III. Clerical				1	8	16	56	38	3		122	115·9
IV. Skilled			2	11	51	101	78	14	1		258	108·2
V. Semiskilled		5	15	31	135	120	17	2			325	97·8
VI. Unskilled	1	18	52	117	53	11	9				261	84·9
Total	1	23	69	160	247	248	162	67	21	2	1000	100·0

TABLE II. DISTRIBUTION OF INTELLIGENCE ACCORDING TO OCCUPATIONAL CLASS: CHILDREN

	50– 60	60– 70	70– 80	80– 90	90– 100	100– 110	110– 120	120– 130	130– 140	140+	Total	Mean I.Q.
I. Higher Professional						1		1	1		3	120·8
II. Lower Professional				1	2	6	12	8	2		31	114·7
III. Clerical			3	8	21	31	35	18	6		122	107·8
IV. Skilled		1	12	33	53	70	59	22	7	1	258	104·6
V. Semiskilled	1	6	23	55	99	85	38	13	5		325	98·9
VI. Unskilled	1	15	32	62	75	54	16	6			261	92·6
Total	2	22	70	159	250	247	160	68	21	1	1000	100·0

But still more striking is the wide range of individual differences within each class. With a normal distribution the range for an unselected group of 1000 cases would be (as in fact appears from the table) nearly 100 I.Q. points—i.e., from about 50 to 150; for an unselected group of 100 cases it would be about 75 to 80 I.Q. points. We should naturally expect, however, that the members of the occupational classes will form selected groups, and that their range and standard deviation will therefore be appreciably diminished. In point of fact the standard deviation within the various classes averages 9·6 (i.e. rather less than two-thirds the standard deviation of the entire group, 15). Hence the range for 100 cases would still be nearly 50 I.Q. points (as a glance at classes III to VI will confirm). Indeed, in the lowest class of all—that of unskilled workers—some of the brightest members actually display greater intelligence than the dullest members in class II, the 'lower professional'. The correlation between intelligence and occupational class therefore is by no means perfect. If we attempt to estimate it on the assumption that both distributions are in fact normal, it works out at just over 0·74. However, since the correlation must be far from linear, its precise numerical value as thus calculated can have little meaning.

The fact that the correlation is far from perfect must not be taken to imply that the duller members of the higher classes and the brighter members of the lower classes are all of necessity instances of vocational misfit. No doubt, they sometimes are. But frequently specific abilities or disabilities, and still more often qualities or infirmities of character and temperament, will fully account for the apparent discrepancies.

The Ideal Redistribution. In order to determine what is the maximum amount of interchange that ideal conditions could possibly permit, let us suppose that vocational adaptation depends solely on intelligence. Then in terms of the I.Q. scale the borderlines between the several occupational classes would be 141, 127, 115, 103, and 90 respectively, and there should be no overlapping between the successive categories. If we now reclassify the actual data for adults according to these new borderlines, we obtain the distribution set out in Table III. The number who are placed in occupations corresponding with their intelligence is shown in semi-bold. Before calculating the percentages let us pool the first three classes together to form a single group which is mainly non-manual; and let us combine the next two to form a group of skilled workers (predominantly but not entirely manual), and leave the lowest class as it is. With this threefold rearrangement (Table V) we find that only 55 per cent of the population could be regarded as correctly placed if intelligence were the sole criterion: nearly 23 per cent are in a class too high, and, with a perfect scheme of vocational selection, ought to be moved down: 22 per cent are in a class too low, and would have to be moved up.

The Distribution for Children. When we turn to the data for children (Table II), we observe that the differences between the class-means are much smaller. The average intelligence of the children in the higher groups has

TABLE III. DISTRIBUTION OF INTELLIGENCE ACCORDING TO OCCUPATIONAL CLASS: ADULTS

Rescaled

	VI 50 – 91	V 91 – 103	IV 103 – 115	III 115 – 127	II 127 – 141	I 141 +	Total
I					2	**1**	3
II			1	15	**14**	1	31
III	1	15	38	**56**	12		122
IV	16	86	**114**	38	4		258
V	53	**178**	84	10			325
VI	**191**	46	21	3			261
Total	261	325	258	122	32	2	1000

TABLE IV. DISTRIBUTION OF INTELLIGENCE ACCORDING TO OCCUPATIONAL CLASS: CHILDREN

Rescaled

	VI 50 – 91	V 91 – 103	IV 103 – 115	III 115 – 127	II 127 – 141	I 141 +	Total
I			1	1	1		3
II	1	4	11	9	**6**		31
III	11	28	51	**20**	12		122
IV	46	66	**75**	62	8	1	258
V	91	**122**	84	23	5		325
VI	**112**	105	36	7	1		261
Total	261	325	258	122	33	1	1000

TABLE V. ADULTS: PERCENTAGE IN EACH GROUP WHOSE INTELLIGENCE IS BELOW, ABOVE, OR EQUIVALENT TO THAT OF THEIR OCCUPATIONAL CLASS

	Below	Equivalent	Above	Number
Class I–III	46·2	45·5	8·3	156
Class IV–V	26·6	50·1	23·3	583
Class VI	—	73·2	26·8	261
Total population	22·7	55·4	21·9	1000

TABLE VI. CHILDREN: PERCENTAGE IN EACH GROUP WHOSE INTELLIGENCE IS BELOW, ABOVE, OR EQUIVALENT TO THAT OF THEIR OCCUPATIONAL CLASS

	Below	Equivalent	Above	Number
Class I–III	75·5	16·8	7·7	156
Class IV–V	34·8	34·3	30·9	583
Class VI	—	42·9	57·1	261
Total population	32·1	33·5	34·4	1000

fallen almost half-way towards the mean of the whole population; similarly that of the children in the lower groups has risen by a similar proportion. There is, in short, an overall regression averaging 0·52 (cf. [1], [3]). The figure is very close to the value we should expect on the assumption that the correlation between fathers and sons was due chiefly to multifactorial inheritance with assortative mating and incomplete dominance. If anything, the coefficient is slightly higher than we might expect on these grounds alone [cf. 14]. Hence environmental influences may perhaps have contributed to increase it; but, if so, the contribution must be extremely small.

The phenomenon just noted has sometimes been termed 'biological regression'; and several sociological and psychological writers have claimed that this tendency is responsible for 'the steady progress which' (so they hold) 'most populations are continually undergoing from a state of individual diversity to one of increasing individual equality', so that in the more highly civilized communities the distribution of intelligence is approaching 'near-randomness' as regards both classes and individuals. This interpretation appears to have been adopted as a corollary to the theory of 'blended inheritance' to which the majority of psychologists and sociologists still adhere[1]. A few writers, however, who recognize that the data obtained from successive generations reveal no evidence whatever for this alleged tendency towards equality, have also postulated a biological 'egression *from* the mean', which, they argue, "balances regression *towards* the mean" [22]. For this there is no need. With multifactorial inheritance this 'conservation of variance' is what we should expect.

When we look at the distribution of children's intelligence *within* the several occupational classes and compare it with that of their fathers' (Table I), we see at once that, so far from progressing towards equality, the amount of individual difference has actually increased. The standard deviation has gone up from 9·6 to 14·0, not far short of the standard deviation for the whole population (15·0). Or, to put it in another way, the range for 100 individuals selected according to occupational class has increased from 50 I.Q. for adults to nearly 75 I.Q. for their children.

One incidental consequence of this increase in variability is the appearance of bright children among the offspring of dull parents in the lower occupational classes and of dull children among the offspring of highly intelligent parents in the upper occupational classes. Consider, for example, the lowest occupational class of all. Among the adults only 20 persons out of 261 have an intelligence above the general average; among the children as many as 76, nearly four times as many—a discrepancy of 56. Dr. Floud, and others who hold as she does that differences in intelligence are due wholly to environmental advantages or disadvantages, can hardly maintain that the high level reached by these 76 boys

---

[1] This corollary from the theory of blended inheritance is mathematically deduced, and empirically disproved, in an earlier issue of this *Journal* (X, p. 56). The empirical disproof of the corollary is perhaps one of the simplest and most convincing arguments against the doctrine of blending. Clarke's 'egression' is apparently a substitute for Darwin's 'spontaneous variation',

—all children of unskilled workers—results from the superior advantages which their home environments confer. But equally, those who adopt the traditional theory of blended inheritance, would find it quite impossible to explain the higher intelligence of these children in terms of their heredity. On the Mendelian hypothesis, however, such apparent anomalies are exactly what we should anticipate if the amount of a child's intelligence is determined mainly, or at any rate largely, by his genetic constitution, and if that in turn is the result of a chance recombination of parental genes ([14], p. 97).

Similar arguments hold good for the marked discrepancies discernible in the upper part of the distribution. In the first three occupational classes, for example, we see that among the adults only 9 out of 156 had an intelligence below the general mean, among the children as many as 39. Here again the increased numbers would be almost inexplicable on the environmental theory, but a natural consequence of the Mendelian theory of polygenic inheritance.

Consequences of the Intergenerational Changes. We have seen that two changes result from the comparatively moderate correlation that obtains between the intelligence of parents and their children: (i) the mean intelligence of the children belonging to each occupational class deviates far less than the mean of the parents from the average for the population as a whole, and (ii) the intelligence of the individual children within any one class varies over a far wider range than that of their parents. Moreover, unless their effects are in some way counteracted, both these changes will be cumulative. After about five generations the differences between the class-means would virtually vanish, and the proportional range within each class would spread out almost as widely as the proportional range of the population as a whole[1].

Now all the evidence shows (p. 5 above) that in point of fact, during the period with which we have been concerned, the occupational distribution of intelligence has remained fairly constant from one generation to the next, and it appears likely to do so in the immediate future. If therefore, when they are grown up, the children of Table IV are to exhibit the same distribution as the adults of Table I, it follows that a considerable number will have to move into a fresh occupational class. Some will go up the social ladder by one or more rungs; others will go down. One of our chief problems therefore is to assess the extent of this migration. For this purpose it will be helpful to begin by rearranging the figures for the children according to the method we have already adopted for the adults (Table III above).

Maximum Mobility. Table IV shows the distribution of the children with the scale for intelligence subdivided afresh so that the lines of division shall correspond with those we should expect between the different occupational classes if occupational efficiency depended solely upon intelligence. As before, let us group together classes I, II and III to form a non-manual group, and classes IV and V together to form a group of skilled workers, leaving class VI as it stands to form a group of unskilled workers. Then, assuming intelligence to be the sole criterion, it appears (as a little mental calculation will quickly show) that in the highest group 75 per cent of the children have an intelligence below the minimum that would be needed if they were to become efficient members of the occupational group into which they were born; on the other hand, in the lowest group 57 per cent have an intelligence well above the meagre amount required for an unskilled worker (see Table VI). In the entire sample over a third of the children have an intelligence which would apparently fit them for a higher occupational class than that of their fathers, and rather less than a third have an intelligence which would be more appropriate for a lower class.

---

[1] This is a simple mathematical corollary. Allowing for assortative mating and partial dominance, the correlation between occupation and intelligence after n generations would sink to approximately $0{\cdot}74 \times \frac{1}{2} \times (2/3)^{n-1}$: (See [14], p. 116, eq. 23).

These figures give a rough indication of the amount of movement upward or downward from one class to another which the children would have to undergo when grown up in order that the type of work they secured corresponded with their degree of intelligence—always assuming that intelligence was the sole criterion. In point of fact we know that nothing like this amount of movement actually takes place; and the figures, of course, merely indicate the *maximum* degree of mobility that is theoretically conceivable.

Actual Mobility. Let us now return to the question of fact. The ideal and most direct procedure would be to plan a longitudinal study of a large and representative sample, following up the children from school to middle life. A complete inquiry of this kind has so far never been attempted. Both in America and in this country follow-up studies have been undertaken for certain selected groups—the gifted or the backward; but for our present problem these provide at most only supplementary or confirmatory evidence. We are obliged therefore to fall back on the alternative procedure commonly adopted in similar situations; and, instead of comparing the same group at two widely diverging intervals of time, we shall compare two different groups of widely divergent ages.

Our present data supply us with two such samples. These are comparable, since the adults are the parents of the children. However, there is a difference of 28·4 years between the average age of the children and the average age of the adults; and, as we have seen, during that amount of time there would have been a variety of changes in the population. Our method of reducing the figures observed to numbers per 1000 should sufficiently allow for the change in the absolute size of the population. The differential birthrate may have entailed some slight modification in the mental quality of the population; but, in the space of three decades only, the extent of the change would, as we have seen, be almost negligible. The effects of the deathrate are largely ruled out by the fact that we have taken children who have survived to school age. Although between 1911 and 1951 the proportion of men and women over 65 very nearly doubled, this was offset by a decline in the number of boys and girls under 15: and the proportional number of males of employable age has remained much the same. The type of work available has changed appreciably: the number of those engaged in manufacturing and in professional and administrative work of various kinds has increased; the number engaged in agriculture, in the extractive industries (mining, quarry, etc.), in domestic work, and in the distributive trades has diminished; moreover, the amount of prestige attaching to different types of occupation has altered. Nevertheless, these further changes are hardly relevant to our present problem, as we have formulated it, although in a more intensive study the bearing of all the varying conditions I have mentioned should undoubtedly be systematically examined.

Assuming then that the data in our two samples are reasonably comparable, our primary task is to determine what kind of compensating change would be necessary to bring the frequency distributions for the children (Table II) into conformity with the frequency distribution for the adults (Table I). Let us look first at the lowest occupational class of all—the unskilled workers (class VI). Among the children, it will be remembered, as many as 57 per cent have an intelligence above what is required for work of this type as against 27 per cent of the adults (Tables V and VI). Hence $(57-27)=30$ per cent of the children will presumably move up to a higher occupational class as they grow up. Similarly $(75-46)=29$ per cent of the upper group—that comprising classes I, II and III—will move down. In the intermediate group—classes IV and V—the changes both upward and downward will be smaller. Thus, as a comparison of the last lines of the two tables suggests, the over-all mobility will be at least $(55-33)=22$ per cent. This figure I regard as indicating the minimum amount of mobility—the amount that is required to maintain what (if I may borrow a phrase from the astronomers) might be called a 'steady state'[1]. It constitutes what may be termed 'basic mobility'.

---

[1] The foregoing data and the analysis I have here attempted will, I hope, dispose of one of the strongest objections urged by Dr. Halsey [21] against the arguments brought forward by Miss Conway and myself in our endeavour to account for the wide differences in average intelligence shown by the different socio-economic classes. Dr. Halsey's criticism was that the round figures assumed for social mobility in setting up our genetic model were far too high. But our object then was of course very different from our present purpose. We merely wanted to show that, with a comparatively small amount of interchange between the several classes, a society which started from primitive conditions in which the average intelligence in the different classes was practically the same would, in the course of subsequent generations, be differentiated in such a way that the differences between the mean levels of intelligence corresponded pretty closely with those at the present day. For this purpose we deliberately assumed in our hypothetical model an amount of mobility well below that which we believed had actually occurred in order to forestall incidental criticisms on this point. In view of the figures given by sociologists themselves Dr. Halsey's criticism seemed rather surprising. But we now hope that the foregoing analysis will show that our postulated figure was well below the most probable minimum.

It has been objected that any figure for social mobility, like that given above, is bound to vary with the lines of division adopted in classifying occupations. However, as long as the basis of the classification remains unaltered, changes in the lines of division will not seriously affect the estimated figure unless the lines of division become so few and the resulting classes so large that the amount of movement is obscured. Indeed, if we imagine the various occupations to be graded according to difficulty in such a way that the distribution of the employees is approximately normal, then in theory, provided we know (i) the correlation between the intelligence of the employees and grade of the occupation, and (ii) the correlation between the intelligence of the employees' children and that of the employees themselves, the amount of mobility required to keep the population constant could be determined from the properties of the bivariate frequency distribution.

However, as we have already seen, there was, at the time when the occupations of the fathers were recorded, considerable room for improvement in the degree of adjustment between the capabilities of the individuals and the type of employment they followed. Moreover, during the last forty years or so, as several researches in the field of vocational guidance have shown, the degree of adjustment has appreciably increased, and apparently is still increasing. Hence the actual amount of mobility is probably much greater than that which just suffices to maintain the *status quo ante*. If we may trust the most thorough of the recent inquiries [13], the overall amount of mobility would appear to be in the neighbourhood of 29 per cent—well above our minimum for a steady state, but still far below what the ideals of vocational suitability would require. Much the same figure was obtained from our analysis of after-histories (p. 19)—viz. 31 per cent.

As the reader will realize, the foregoing deductions deal only with a very limited aspect in a very limited interpretation of the rather ambiguous phrase 'social mobility'. Ordinarily the discussion of mobility treats a rise in social status as implying something more than a mere rise to a type of occupation which requires a higher I.Q. than the occupation followed by one's father[1]; and the value attached to different types of occupation as a goal for the ambitious youngster varies widely from group to group, from individual to individual, and from one period to another. Nor is intelligence the only factor which determines whether or not the ambitious youngster will succeed in achieving the vocational career at which he aims.

I propose therefore in conclusion to glance at two or three other factors which might be expected to influence the kind of occupational status which persons of varying intelligence are likely to attain, and consider what is their relative importance and how far they could affect the inferences already drawn.

V. CAUSAL FACTORS

Data. Although opinions have been freely expressed about the conditions which facilitate social advancement and still more often about those which are thought to obstruct it, surprisingly little factual evidence has been obtained. The teaching of Samuel Smiles and his Victorian doctrine of 'Self-Help' has long since faded from memory, though his biographical illustrations are by no means valueless. However, during the past fifty years the popular tendency has been to place an increasing emphasis on the external or social factors and less on the personal or psychological. But here it is principally the latter which I should like quite briefly to examine.

---

[1] As I have pointed out elsewhere, in their definitions of social class and social status different writers have relied on a wide variety of criteria ([12], pp. 37f. and refs.). The problems to which discussions of this type give rise are most readily handled by means of factor analysis. It turns out that all the criteria are positively, and indeed closely, correlated: so that a general factor must underlie them all. Hence the ideal way of allocating persons to an appropriate social class would be to use a system of weights based on a multiple regression equation. The use of such a technique of course implies that the necessary data have been collected for all the individuals concerned.

As part of the longitudinal studies of gifted and of backward children attending London County Council schools my colleagues and I have followed up into later life a large number of cases, not only of these somewhat exceptional types, but also of normal or average children who were treated as control groups. We now possess fairly detailed data for just over 200 ordinary children who have already reached an age when it is possible to say either that they have already moved out of their original class, in one direction or the other, or else that it is now practically certain that they will never do so. We have similar numbers for pupils who formerly attended central schools or won junior county scholarships as well as for pupils who were educationally subnormal. By using fractional weights for the figures obtained from these various subgroups we can compile a composite group of males which shall be reasonably representative of the total population. It includes many of the older children in the group discussed in section II—those whom we have been able to trace and follow up in their after-school life; but it includes others omitted from that group owing to lack of adequate data about their parents' abilities. For each of the sub-groups we have the following relevant information, obtained (except for vi) mainly when the children were at school: (i) the occupational class of the fathers at the time the children were born; (ii) assessments and descriptions of the home background, and particularly of the attitude of the family towards the child's social advancement; (iii) the child's own attitude, and particularly his industry, ambition, and educational and vocational aims; (iv) his level of intelligence, based on tests duly checked with the teachers and corrected where necessary; (v) his educational record (more especially his admission to a grammar school or its equivalent); (vi) his occupation when last visited.

In view of the complexity of the problem and the limitations of the data let us begin by an analysis in terms of crude percentages. For this purpose we may divide the whole composite group into three portions—(*a*) those who have remained in their original occupational class, (*b*) those who have moved up, and (*c*) those who have moved down. Similarly, it will simplify matters to reduce the assessments for intelligence to a threefold classification—(*a*) an intelligence equal to that required in the individual's original occupational class (the class of his father), (*b*) an intelligence above it, and (*c*) an intelligence below it. The other assessments can be reduced to a twofold classification, viz. (*a*) above the median and (*b*) below the median[1].

Results. The main results are shown in Table VII. The following conclusions may be drawn.

1. Of the children with an intelligence *below* the minimum required for the occupational class into which they were born none rose above it, and about a third (or rather more) dropped to a lower class. On the other hand, of those

---

[1] For much of the data in this section, and most of the calculations, I am deeply indebted to my former colleagues, Miss J. L. Hastings, Miss E. Davenport, and Mr. R. M. Weldon.

who had an intelligence *above* the maximum required for their original occupational class nearly 60 per cent failed to rise.

2. Very few of those who were assessed as lacking in adequate motivation rose above their original class. Indeed, poor motivation was more likely than poor intelligence to contribute to a fall. On the other hand, good motivation was less certain to secure a rise.

3. A good home background, though helpful particularly during the earlier educational stages, was less effective in securing a rise than either high intelligence or strong motivation. Nor was a bad home background so fatal as seems to be commonly assumed. Nearly a quarter (24 per cent) of those who suffered from unfavourable home circumstances in childhood nevertheless succeeded in rising out of their original class.

TABLE VII. PSYCHOLOGICAL FACTORS INFLUENCING OCCUPATIONAL MOBILITY

	Intelligence			Motivation		Home Background		Educational Achievement	
Mobility	P	A	G	P	G	P	G	P	G
Up	0	12	41	2	36	24	29	18	34
Stationary	64	67	49	47	51	40	44	34	52
Down	36	21	10	51	13	36	27	48	14
	100	100	100	100	100	100	100	100	100

	Intelligence and Motivation Combined		Intelligence, Motivation, and Home Background Combined		Intelligence, Motivation, Home Background, and Educational Achievement Combined	
Mobility	P	G	P	G	P	G
Up	0	66	0	70	0	72
Stationary	32	29	23	23	18	21
Down	68	5	77	7	82	7
	100	100	100	100	100	100

Note.—P=Poor, A=Average, G=Good. The tables for 'combined' qualities include only those cases in which *all* the qualities specified were 'Poor' or 'Good'.

4. The achievement of grammar school status or (during the pre-war period) the award of a junior county scholarship, by no means sufficed to guarantee a rise in occupational class, though it often proved an important step in the child's gradual ascent. However, an appreciable number succeeded in working their way up to a higher occupational level in spite of a total lack of any formal education beyond what the elementary school could provide.

5. Two-thirds of those who have both high intelligence and strong motivation are likely to achieve an occupational rise. In fact nearly all who achieve a rise have this double characteristic. The addition of a good home background

is a further advantage; in the case of the child of the lower classes what chiefly count are the social aspirations, the ambitious aims, and the constant urging that often characterize the more earnest working-class parents; with children from higher levels it is rather the intellectual and cultural character of the home that helps. The addition of a grammar school education does not greatly increase predictability[1]: this is because most children who have good intelligence, good motivation, and good home backgrounds are pretty sure to win their way to a grammar school. Those who, despite high intelligence and good motivation, dropped to a lower occupational class were for the most part victims of ill health, either mental or physical. As the tables indicate, there was actually an increase in this type of failure among those who in addition enjoyed a good home and a good education; the increase, as the case-histories would show, is accounted for by the larger number who break down from nervous ill health.

There are many other possible factors of a somewhat miscellaneous kind which have not been included in the foregoing summary—e.g. the variations in the openings available at different times or in different regions, the effects of a wife and family, or of friends, acquaintances, or patrons able to help and wielding personal influence. From time to time we encountered evidence of such factors; but as determinants of individual mobility they seemed to be of much less importance than those we have discussed.

An Analysis by Factors and Variances. The relations between the five main variables can be roughly expressed by means of correlation coefficients. For this purpose we have assessed mobility in terms of the degree of movement as well as the direction, allotting ± 1 point for a movement to the class above or below, ± 2 points for a movement over two classes, and so on. The correlations obtained are shown in Table VIII. The correlations between social mobility and the various causal conditions are product-moment coefficients; the correlations of the causal conditions with each other are averaged tetrachorics.

The raw correlations between social mobility and the four main causal conditions differ but little in magnitude; but the partial correlations (last row of Table VIII) differ appreciably. The correlation with intelligence is by far the highest of the three (0·38); the correlation with motivation is somewhat smaller (0·29); and the correlations with home background (0·17) and educational achievement (0·05) almost negligible. These figures thus fully confirm the conclusions already reached. The partial regression equation for predicting mobility is

$$S = 0{\cdot}346I + 0{\cdot}272M + 0{\cdot}158H + 0{\cdot}149E,$$

where S denotes social mobility, I intelligence, M motivation, H home background, and E educational achievement. The multiple correlation is 0·628.

---

[1] Education, and particularly the type of education (e.g. entrance at a public school followed by Oxford or Cambridge), are unquestionably influential in determining a rise at the highest levels of all; but in a survey of the total population such cases appear by comparison few in number and exceptional in type.

This rather modest value is no doubt due to the fact that, as we have seen, various minor factors may be operative in individual cases[1].

Table IX gives the results of the factor analysis. The matrix of correlations between the four causal conditions was first subjected to a group factor analysis, the lines of division between the groups being determined by a preliminary bipolar analysis. The correlations between social mobility, etc., and the resulting factors were then computed by the usual formula. In determining the meaning of the factors we have relied partly on the case-histories of a few typical individuals who have obtained exceptionally high or exceptionally low factor measurements for each of the factors. The first factor appears to be essentially an intellectual factor; and the second a factor of incentive. The last two are based on a bipolar factor which apparently distinguishes variations that are mainly genetic from variations that are mainly environmental. But the precise interpretation of factor III remains somewhat obscure[2].

TABLE VIII. CORRELATIONS BETWEEN SOCIAL MOBILITY AND RELATED CONDITIONS

	1	2	3	4	5
1. Social Mobility	—	0·481	0·402	0·396	0·378
2. Intelligence	0·481	—	0·133	0·236	0·413
3. Motivation	0·402	0·133	—	0·376	0·162
4. Home Conditions	0·396	0·236	0·376	—	0·363
5. Educational Record	0·378	0·413	0·162	0·363	—
Social Mobility (partial correlation)	—	0·379	0·286	0·174	0·047

Note.—The partial correlations give the correlation between social mobility and the condition specified when the effects of the other three conditions are held constant.

TABLE IX. CORRELATIONS OF SOCIAL MOBILITY AND RELATED CONDITIONS WITH GROUP FACTORS

Factor	I	II	III	IV
1. Social Mobility	0·443	0·417	0·294	0·166
2. Intelligence	0·782	0·000	0·405	0·000
3. Motivation	0·000	0·731	0·329	0·000
4. Home Conditions	0·302	0·514	0·000	0·287
5. Educational Record	0·528	0·216	0·000	0·323

---

[1] The method is similar to that adopted in an earlier memorandum and subjected to considerable criticism: for a reply see [6], pp. 278f.

[2] Factors I and II must each of them be partly determined by genetic characteristics. Factor III seems to imply some overlap or linkage between the genetic characteristics that make for intelligence and stability of character respectively—perhaps due merely to the fact that intelligence is an ingredient of stability, or perhaps due to some selective conditions influencing the genetic basis of both.

Between them the four factors contribute about 48 per cent of the total variance for mobility: of this, factor I contributes nearly 20 per cent, factor II 17 per cent, factor III 9 per cent, and factor IV barely 3 per cent.

Owing to the imperfect nature of the data and the methods of calculating the correlations it seems very doubtful whether much value can be attached to the figures thus obtained. We give them chiefly in the hope that their shortcomings may encourage fresh investigators to plan a more systematic set of longitudinal studies, with the method of analysis kept carefully in view from the very outset, and so in the end reach a more reliable basis of comparison[1].

VI. SUMMARY AND CONCLUSIONS

1. As a convenient criterion for vocational adjustment it is assumed that, if the available occupations are grouped in order of the difficulty of the work they entail, and if the men engaged on them are grouped in order of intelligence, then there should be a perfect correspondence between the two series. Judged by this criterion it appears that well over 20 per cent of the male adults in this country have a higher intelligence than is requisite for the work they are doing and that about the same number have an intelligence which is inadequate. Many of the discrepancies, however, are accounted for by individual differences in qualities of personality or character or in special abilities or aptitudes relevant to the work concerned.

2. Owing to the imperfect correlation between the intelligence of parents and the intelligence of their children the discrepancies between the children's intelligence and the occupational category of the parents are still greater. This follows from the multifactorial theory of inheritance, and is amply confirmed by the data here examined. The figures indicate that an overall mobility of about 22 per cent is needed to keep the distribution of intelligence approximately constant from one generation to another within each occupational category. If the distribution of character-qualities could also be taken into account, a still higher figure would no doubt be obtained. There is, moreover, considerable evidence to suggest that the degree of general vocational adaptation is improving; and, partly for this reason, it is estimated that the total amount of intergenerational mobility must be nearer 30 per cent.

3. Of the various causal factors affecting the individual's rise or fall in occupational status differences in intelligence and motivation appear to be the most influential. Differences in home background and in education seem to exercise a secondary or supplementary influence, but without the basis of the first two they are of little effect.

---

[1] I am much indebted to Miss Howard for assistance with the calculations involved in this section.

REFERENCES

[1] BURT, C., SMITH, M., *et al.* (1926). *A Study in Vocational Guidance.* London: H.M. Stationery Office.
[2] CARR-SAUNDERS, A. M. and JONES, D. C. (1937). *Social Structure in England and Wales.* Oxford: Clarendon Press.
[3] BURT, C. (1942). Ability and income. *Brit. J. Educ. Psychol.*, XIII, 83–98.
[4] BURT, C. (1943). *The Backward Child.* London: University of London Press.
[5] BURT, C. (1946). *Intelligence and Fertility.* London: Hamish Hamilton.
[6] BURT, C. (1947). *Mental and Scholastic Tests.* 2nd ed. London: Staples Press.
[7] Scottish Council for Research in Education (1949). *The Trend of Scottish Intelligence.* London: University of London Press.
[8] BURT, C. (1950). The trend of Scottish intelligence. *Brit. J. Educ. Psychol.*, XX, 55–61.
[9] LIVI, L. (1950). Sur la mesure de la mobilité sociale. *Population*, V, 65–76.
[10] Registrar-General (1951). *Census of England and Wales: 1951.* London: H.M. Stationery Office.
[11] Registrar-General (1951). *Decennial Supplement: England and Wales.* London: H.M. Stationery Office.
[12] BURT, C. (1953). *Contributions of Psychology to Social Problems.* London: Oxford University Press.
[13] GLASS, D. V. (1954). *Social Mobility in Britain.* London: Routledge and Kegan Paul.
[14] BURT, C. and HOWARD, M. (1956). The multifactorial theory of inheritance and its application to intelligence. *Brit. J. Statist. Psychol.*, IX, 95–131.
[15] TULDER, J. J. M. VAN, (1956). Occupational mobility in the Netherlands. *Trans. World Congr. Sociol.*, III, 209–218.
[16] FLOUD, J. E. and HALSEY, A. H. (1956). *Social Class and Educational Opportunity.* London: Heinemann.
[17] Japanese Sociological Research Committee (1957). *Social Stratification and Social Mobility.* Tokyo: Japanese Sociological Association.
[18] JANOWITZ, M. (1958). Social stratification and mobility in Western Germany. *Amer. J. Sociol.*, LXIV, 6–24.
[19] BURT, C. (1959). Class differences in general intelligence. *Brit. J. Statist. Psychol.*, XII, 15–34.
[20] CONWAY, J. (1959). Class differences in general intelligence. *Brit. J. Statist. Psychol.*, XII, 5–14.
[21] HALSEY, A. H. (1959). Class differences in general intelligence. *Brit. J. Statist. Psychol.*, XII, 1–4.
[22] CLARKE, A. D. B., CLARKE, A. M., and BROWN, R. I. (1960). Regression to the mean. *Brit. J. Psychol.*, LI, 105–118.
[23] LIPSET, S. M. and BENDIX, R. (1960). *Social Mobility in Industrial Society.* London: Heinemann.
[24] Registrar-General (1960). *Statistical Review for 1959.* London: H.M. Stationery Office.
[25] FLOUD, J. and HALSEY, A. H. (1961). Homes and schools: social determinants of educability. *Educ. Research*, III, 83–86.
[26] BURT, C. (1961). The gifted child. *The Yearbook of Education.* London: Evans. (In the press).

Vol. XIV
Part 1
The British Journal of Statistical Psychology
May
1961

SUBJECTIVE DISCRIMINATION AS A STATISTICAL METHOD

By M. Stone

Medical Research Council Applied Psychology Unit

The following paper describes the use of human discriminators to test the hypothesis that there is no difference between two groups of items. Four methods of making a test on the basis of such discriminations are discussed; and the procedure is applied to experimental data.

I. Introduction

The use of statistical tests in psychology has the disadvantage that, for their significance levels to be meaningful, the procedures must be defined before the data are examined. This implies that the hypotheses to be tested must be formulated in advance; accidental discoveries are suspect. The difficulty is overcome by a two-stage procedure. Data from a pilot experiment (or a randomly chosen part of the whole data) are examined. An explanatory hypothesis is then exactly specified, and an appropriate test of it is made in the main experiment (or in the previously unexamined part of the data).

A method which integrates these two stages is that of 'subjective discrimination'. This method will be discussed, and applied to the data of Table I. The table gives five successive decrements in performance (expressed in terms of the percentage of missed signals) at a vigilance task carried out by two groups of subjects, namely, (i) $S_1, \ldots, S_6$, who started the task at 10 a.m. and (ii) $S_7, \ldots, S_{12}$, who started it at 2 p.m. It is of psychological interest to know whether there is any significant difference in the pattern of decrements between the two groups.

Table I. Performance Decrements (D) at a Vigilance Task: G_1 *at* 10 *a.m.*, G_2 *at* 2 *p.m.*

		D_1	D_2	D_3	D_4	D_5
G_1	S_1	3	2	-6	6	-4
	S_2	12	10	-2	13	-4
	S_3	8	-7	-4	4	-3
	S_4	-5	-1	6	15	-3
	S_5	3	3	2	-19	-7
	S_6	8	7	-5	0	-16
G_2	S_7	9	-5	1	0	6
	S_8	15	2	0	-4	17
	S_9	38	-7	-1	-6	2
	S_{10}	26	-6	11	-5	6
	S_{11}	15	2	3	-6	-4
	S_{12}	12	9	-1	0	4

II. THE METHOD

To maintain some generality let us suppose we have two groups of items, G_1 and G_2, m and n in number respectively. (For the data of Table I the items will be graphs of the decrement scores, one for each subject.) The $m+n$ items are presented in a random array to each of J judges, who are the 'subjective discriminators'. They are asked to sort the items into two groups, m and n in number, so that members of the same group shall be more similar than members of different groups. The groupings are recorded. The choice of a similarity criterion is left to the judges, but in most cases it would clearly be useful to ask them to describe their criteria.

III. GENERAL ANALYSIS

Denote the items in G_1 by $I_1, \ldots, I_m$ and those in G_2 by $I_{m+1}, \ldots, I_{m+n}$. Denote the jth judge's m-group and n-group by $J_m{}^j$ and $J_n{}^j$ respectively.

(*a*) *Single judge.* A test of the null hypothesis (that the judge is unable to discriminate between G_1 and G_2) is provided by testing association between them and the judge's groups.

(*b*) *More than one judge.* *Method I* (applicable only when $m \neq n$). Let a_i be the number of judges who place I_i in their n-groups. The null hypothesis that G_1 and G_2 are indistinguishable means that, in independent repetitions of the experiment (or observation) which gave rise to G_1 and G_2, the numbers $a_1, a_2, \ldots, a_{m+n}$ would have a symmetrical distribution with respect to the suffices $1, 2, \ldots, m+n$, that is, the numbers would be distributed in a manner unrelated to the division into G_1 and G_2. The alternative hypothesis of discriminability of G_1 and G_2 would result in the numbers $a_1, \ldots, a_m$ being smaller than $a_{m+1}, \ldots, a_{m+n}$. Therefore suitable tests of the null hypothesis are provided by (i) a randomization test of the difference between the sets $\{a_1, \ldots, a_m\}$ and $\{a_{m+1}, \ldots, a_{m+n}\}$, using $a_1 + \ldots + a_m$ as the test criterion, or (ii) a randomization test with the a_i replaced by their ranks, which is equivalent to the Mann–Whitney rank test.

Method II (a modification of Method I for the case $m=n$). When $m=n$, the decision as to which group shall be denoted by $J_m{}^j$ and which by $J_n{}^j$ is arbitrary. One of the items, say I_s, is selected by the experimenter as a standard; and, for each j, $J_m{}^j$ is defined to include I_s. With the arbitrariness thus removed the numbers a_i are calculated as for Method I (now $a_s = 0$). Provided the selection of I_s is made on a criterion independent of the division into G_1 and G_2, the numbers $\{a_i / i \neq s\}$ will, on the null hypothesis, have a symmetrical distribution: so an appropriate test of the null hypothesis is the same as for Method I, except that the two sets of numbers are now $\{a_1, \ldots, a_m\} - \{a_s\}$ and $\{a_{m+1}, \ldots, a_{m+n}\} - \{a_s\}$. It is clear that, if a standard were chosen about which judgements are erratic, the test would be weakened. To avoid this, s may be chosen so as to maximize

$$\max_{i \neq s} a_i - \min_{i \neq s} a_i.$$

If more than one value of s maximizes this expression, the one used may be chosen from these at random. This criterion for the choice of s satisfies the condition of being independent of the division into G_1 and G_2, and maximizes the agreement between judges in a particular sense. No doubt other reasonable criteria also exist.

Method III (m and n unrestricted). Let a_{ij} be the number of judges putting I_i and I_j in the same group. The null hypothesis that G_1 and G_2 are indistinguishable means that, in independent repetitions of the experiment (or observation) which gave rise to G_1 and G_2, the numbers $\{a_{ij} | i=1, \ldots, m+n; j=1, \ldots, m+n; \; i \neq j\}$ should have a symmetrical distribution with respect to the suffices, that is, they should be distributed in a manner unrelated to the division into G_1 and G_2. There are $\frac{1}{2}(m+n)\;(m+n-l)$ different pairs of curves of which $\frac{1}{2}m(m-1)+\frac{1}{2}n(n-1)$ are from curves in the same group, G_1 or G_2. Under the alternative hypothesis of discriminability, we would expect the a_{ij}'s from the latter pairs to be larger than the a_{ij}'s from the remaining pairs. The following randomization test is therefore appropriate for testing the null hypothesis.

Let $G_1{}^* - G_2{}^*$ denote one of the $(m+n)!/m!n!$ ways of conceptually dividing the $m+n$ items into two groups of sizes m and n. Let S^* be defined by

$$S^* = \sum_{\substack{I_i \in G_1{}^* \\ I_j \in G_2{}^*}} a_{ij}.$$

Then the significance probability is the proportion of the S^*'s which equal or are less than the S^* for the division into G_1 and G_2. To assist the calculation of S^*, the following relation may be used:

$$S^* = \sum_{j=1}^{J} \{b_j{}^*(m-b_j{}^*) + b_j{}^*(n-b_j{}^*)\}$$

where $b_j{}^*$ is the number of items in $G_1{}^*$ and $J_n{}^j$ simultaneously.

Method IV (m and n unrestricted). Proceed as for Method III but use

$$S^* = -\sum_{j=1}^{J} \{b_j{}^{*2} - (m-b_j{}^*)(n-b_j{}^*)\}^2$$

$$= -\sum_{j=1}^{J} \{mn - (m+n)b_j{}^*\}^2.$$

Here S^* is proportional to the sum of J chi-squared statistics for testing the association between $G_1{}^* - G_2{}^*$ and $J_m{}^j - J_n{}^j$ $(j=1, \ldots, J)$.

Comparison of the Methods. Method I is computationally the simplest, and is easy to apply even when m and n are large. However, the work required to select the standard item in Method II is not very great. Methods III and IV would usually require the assistance of an electronic computer, the exception being when all the judges discriminate nearly perfectly. For all the methods there is a limit to the significance level as the number of judges is increased. The limits are $m!n!/(m+n)!$ for Method I, $m!(m-1)!/(2m-1)!$ for Method II,

$m!n!/(m+n)!$ for Methods III and IV when $m \neq n$, $2.m!m!/(2m)!$ for Methods III and IV when $m=n$. In practice the significance level may not increase as J increases; one good judge may do better than a group of judges.

No power comparisons will be attempted here. There is a twofold difficulty in such comparisons. First, it is not easy to specify a sensible alternative hypothesis; secondly, the calculations would be complex. However, it is clear that Method I will not be powerful when $(m-n)/m$ is small and positive, and the judges have probability near $\frac{1}{2}$ of putting G_2 in their m-group.

IV. Application to Experimental Data

Twelve graphs of the data (one for each subject) were presented in random array to each of nine judges. The groupings made are recorded in Table II, where they are distinguished by 0's and 1's. On applying Method II for simplicity, it can be seen that the graph of S_1 is a suitable standard. The Mann–Whitney test yields a significance probability of less than 0·01.

Table II. Results for the Nine Judges

		S.H.	J.L.	H.D.	A.B.	R.W.	E.P.	M.S.	I.H.	I.B.
G_1	S_1	0	0	0	0	0	0	0	0	0
	S_2	0	0	0	0	1	0	0	0	0
	S_3	0	0	0	0	0	0	0	0	0
	S_4	1	0	0	0	1	0	0	0	1
	S_5	1	1	0	1	1	1	0	0	1
	S_6	0	0	0	0	0	0	0	0	0
G_2	S_7	1	1	1	1	0	1	1	1	1
	S_8	0	1	1	1	1	1	1	1	0
	S_9	1	1	1	1	1	1	1	1	1
	S_{10}	1	1	1	1	1	1	1	1	1
	S_{11}	1	1	1	1	0	1	1	1	1
	S_{12}	0	0	1	0	0	0	1	1	0

Alternative methods of analysis do not achieve the same level of significance. With the assumption of multivariate-normality and equality of covariance matrices over the population of experimental subjects, Hotelling's T^2 can be used to test the null hypothesis that G_1 and G_2 have the same means. Using the method of Sect. 5.3.3 of Anderson's *Introduction to Multivariate Analysis*, the differences prove non-significant even at the 5 per cent level. The most outstanding *single* feature of the data is the difference between the two groups under D_1. With a two-tailed Mann–Whitney test this is significant at $P=0{\cdot}008$. However, allowance must be made for the way the feature to be tested is selected. We could have compared any of the five decrements; so that a more realistic probability is $5 \times 0{\cdot}008 = 0{\cdot}04$, since some selection still remains, perhaps even higher.

Vol. XIV
Part I
The British Journal of Statistical Psychology
May
1961

FACTOR ANALYSIS, LATENT STRUCTURE ANALYSIS AND MENTAL TYPOLOGY

By Curtis R. Miller, Richard K. Eyman and Harvey F. Dingman
Pacific State Hospital, California[1]

The following paper deals with the classification of mentally retarded patients into groups or types, and is based on an analysis of data for over 700 cases. Two alternative methods have been used: first, latent structure analysis (a relatively new procedure suggested by Lazarsfeld) and factor analysis (a standard procedure for classifying persons as well as analysing tests or characteristics). It is found that the former yields a classification that agrees more closely with that proposed by earlier investigators in this field on the basis of subjective or intuitive judgement.

I. Problem and Methods

In the field of mental retardation, typologies and diagnostic groupings are as numerous as in other areas of clinical psychology. The number of somatic categories is ever increasing; and psychological categories show no signs of diminishing. Categories of mental retardation have been proposed by Sarason [8], and additional categories discussed by Spitz [9], Strauss [10], and others. Since the early work of Patterson and Rundquist [6] it has been apparent that mentally retarded individuals can be divided into at least two well defined groups on the basis of sociological variables [2, 7, 11]; and, regardless of the particular classifications eventually reached, it has been subjectively felt that there are in fact at least two major types: but to demonstrate such a typology empirically requires much more than subjective evidence.

Latent structure analysis, as proposed by Lazarsfeld [5], was developed so that a population could be divided, on the basis of its manifest characteristics, into 'latent groups'. Each separate group would evidence homogeneity with respect to the manifest characteristics ; but different patterns of characteristics would be found between groups. The development of latent structure analysis depends heavily on the mathematical techniques of multiple factor analysis presented by Thurstone and others [12]. Hence one may ask if either of these methods will confirm the existence of a typology that everyone subjectively feels must exist, and if so, are there any differences between the typologies developed by the two methods ?

The clinical records of all first admissions to Pacific State Hospital during the fiscal years 1947–1952 were examined for information concerning their families, early childhood characteristics, and their present adjustment and capabilities. This information was then punched on IBM cards for the purposes of tabulation. Training of the clerks proceeded to the point that two clerks would agree on four records out of five with respect to every item used in the analysis.

---

[1] Supported in part by the National Institute of Mental Health Grant No. 3M–9130: Population Movement of Mental Defectives and Related Physical, Behavioral, Social, and Cultural Factors. Pacific State Hospital, Pomona, California.

II. Results

Seven hundred and nine patients were admitted to Pacific State Hospital during 1948–1952 with the understanding that they would remain in the institution for treatment. Eight variables were selected from the 195 available from the code sheet. A matrix of joint proportions was computed and is presented in Table I. The variables were analysed on the basis of psychologically and medically meaningful categories. A secondary consideration was given to the distribution of the variables so that no dichotomy would be greater than 70 : 30. One of the two resulting categories was arbitrarily designated as a positive value for the variable.

The matrix of intercorrelations of the eight variables is shown in Table II. A latent structure analysis was performed on the joint proportion matrix, and the data rotated according to the suggestions of Gibson [3]. The rotated loadings are shown in Table III. A factor analysis was carried out with the same data, based on the appropriate correlation coefficients and a complete centroid procedure as described by Thurstone [12]. The factors were rotated to simple structure. The rotated factor loadings are given in Table IV.

Table III reveals that four classes emerge from the latent structure analysis: (in this table, the column at the extreme right indicates the combination of positive responses finally selected for rotation). It might be mentioned that the proper designation of responses as positive facilitates the ease with which rotation may be performed. Since probabilities cannot be greater than 1 or negative, the rotational restriction is often severe [3]. Some combinations of positive responses prove difficult to rotate. Fortunately the unrotated matrix may be altered so as to rearrange positive designations without the necessity of re-factoring [4]

The first latent class is characterized by severely retarded Caucasian children with low IQ's and a somatic diagnosis, whose parents are of a higher socio-economic status. This class corresponds to the severely retarded young group discussed by other authors. The other three classes consist of persons above nine years of age of varying types with higher IQ's. Class II seems to consist of delinquent non-Caucasians, Class III might be characterized as Caucasians with a diagnosis of undifferentiated or familial disorder with high IQ's, and Class IV as older outgoing individuals with high IQ's. It is apparent that Classes II, III, and IV are similar, and might be grouped together into an older, high IQ, delinquent class. Such a grouping would conform to the second major type suggested by various authors.

The factor analysis results were rotated to simple structure, the usual criterion for factor analytic rotations. It must be remembered in interpreting the results of the two methods that a high positive correlation is similar to a high

TABLE I. MATRIX OF JOINT PROPORTIONS

	0	IQ	Release	Race	Diagnosis	Occupation	Delinquency	Direction of Personality	Age	Positive Responses
0		0·391	0·369	0·757	0·564	0·529	0·515	0·589	0·606	
IQ	0·391		0·233	0·233	0·105	0·162	0·202	0·305	0·333	50 +
Release	0·369	0·233		0·203	0·111	0·146	0·115	0·285	0·301	Yes
Race	0·757	0·223	0·203		0·492	0·491	0·436	0·412	0·404	White
Diagnosis	0·564	0·105	0·111	0·492		0·360	0·376	0·251	0·250	Medical
Occupation	0·529	0·162	0·146	0·491	0·360		0·332	0·278	0·269	Skilled and above
Delinquency	0·515	0·202	0·115	0·436	0·376	0·332		0·240	0·240	No
Direction of personality	0·589	0·305	0·285	0·412	0·251	0·278	0·240		0·457	Outgoing
Age	0·606	0·333	0·301	0·404	0·250	0·269	0·240	0·457		Over 9

TABLE II. MATRIX OF CORRELATION COEFFICIENTS

	IQ	Release	Race	Diagnosis	Occupation	Delinquency	Direction of Personality	Age	Positive Responses
IQ		0·57	0·25	0·44	0·20	0·55	0·40	0·75	50 +
Release	0·57		0·31	0·39	0·25	0·42	0·32	0·44	Yes
Race	0·25	0·31		0·30	0·40	0·26	0·19	0·28	White
Diagnosis	0·44	0·39	0·30		0·26	0·39	0·27	0·42	Medical
Occupation	0·20	0·25	0·40	0·26		0·45	0·35	0·30	Skilled and above
Delinquency	0·55	0·42	0·26	0·39	0·45		0·33	0·50	No
Direction of personality	0·40	0·32	0·19	0·27	0·35	0·33		0·51	Outgoing
Age	0·75	0·44	0·28	0·42	0·30	0·50	0·51		Over 9

NOTE.—The correlation between Race and Diagnosis is the Phi-Coefficient ; the other correlations for these two variables are Point Biserial Coefficients. The remaining correlations are Pearson Product Moment Coefficients.

latent class proportion, but a latent class proportion of 0·50 is similar to a factor loading of zero and a latent class proportion of zero implies a high negative correlation. Comparing the results, Factor I is similar to latent Class I and Factor II is somewhat like latent Classes II and IV, but further identification is impossible.

It is interesting that the factor that gives the greatest loading on release rate gives the highest loading on IQ, where, in the latent structure analysis, the highest release rate was found in a group with rather low IQ's. Another anomalous finding is in rotated Factor II, where we find that non-delinquency and high socio-economic status are loaded on a factor where the loading for Caucasian is only 0·25. This is inconsistent with the latent structure analysis but not with the results published by Sabagh wherein he examines the same data according to incidence rates.

Table III. Rotated Group Structure

Variable	Class I	Class II	Class III	Class IV	Positive Responses
0	1·00	1·00	1·00	1·00	
IQ	0·06	0·40	0·85	0·89	50 +
Release	0·92	0·26	0·54	0·35	No
Race	0·01	0·81	0·12	0·34	Non-white
Diagnosis	0·91	0·24	0·00	0·33	Medical
Occupation	0·27	0·99	0·65	0·42	Low
Delinquency	0·74	0·72	0·29	0·54	None
Direction of personality	0·68	0·30	0·18	0·02	Withdrawn
Age	0·71	0·20	0·09	0·01	Under 10
Proportion in each class	0·49	0·18	0·09	0·23	

Table IV. Rotated Factor Loadings

	I	II	III	IV	Positive Responses
IQ	−0·85	0·14	0·24	0·08	50 +
Release	0·62	−0·11	−0·09	−0·30	No
Race	−0·13	0·25	0·20	0·50	Non-white
Diagnosis	0·44	−0·18	−0·09	−0·36	Medical
Occupation	−0·02	0·56	0·30	0·37	Medical
Delinquency	0·51	−0·61	−0·05	−0·15	None
Direction of personality	0·33	−0·20	−0·52	−0·10	withdrawn
Age	0·65	−0·23	−0·53	−0·07	Under 10

III. A Comparison of the Two Methods

Latent structure analysis, a method for dividing a population into categories based on manifest characteristics, appears essentially to substantiate the existence of a typology which is well recognized in the field of mental retardation. A factor analysis of the same data does not seem to provide such a simple and coherent picture. One reason for this finding could well be associated with assumptions of the two techniques as discussed by Gibson [3], Green [4], and Lazarsfeld [5]. The latent structure equations do not imply any assumptions concerning dimensionality or linearity. It is simply a method for extracting 'latent classes' within which responses to items are independent in the coin-tossing sense. Factor analysis, which is based on correlations, implies linearity between tests and factors, and is concerned directly with dimensionality. If all the assumptions of factor analysis were met, the correspondence between the models would be very close (Gibson [3]); but differences arise when the necessary conditions are not fully met. In addition, it is also possible that a different rotational solution would produce greater clarity in the factor structure. The solution given essentially followed the common rules of factor analytic rotation, and did not attempt to rotate to maximum agreement between the two methods.

A second hypothesis is that the latent class analysis is less sensitive to variations within groups, and would therefore tend to produce gross typologies. Furthermore there is the possibility that the factor analysis reflects more of the variation within sub-classes and that the latent structure results reflect groups of patients within broad categories.

A final comment must take note of the type of correlations involved in the factor analysis. An examination of the variables demonstrates that three types of correlations (Pearsonian *r*, point biserials, and phi) are present in Table III, the matrix of intercorrelations. The effect of this mixing is unknown; but in view of the discussion in Dingman [1] this should not be discounted.

IV. CONCLUSIONS

A relatively new procedure proposed by Lazarsfeld seems to produce a typology that is subjectively accurate as judged by the views of numerous investigators in the field of mental retardation. A similar technique in common use for the same purpose is factor analysis. This makes more assumptions; but, although the two statistical procedures are closely analogous, factor analysis here yields a picture which is not entirely consistent with that presented by the latent structure analysis. No doubt factor analysis and latent structure analysis have each its own particular rôle in typological investigations; but only further research can indicate under what specific conditions either is the more appropriate.

REFERENCES

[1] DINGMAN, H. F. (1958). The relation between coefficients of correlation and difficulty factors. *Brit. J. statist. Psychol.*, XI, 13–17.

[2] DINGMAN, H. F. (1961). Some uses of descriptive statistics in population analysis. *Amer. J. ment. Def.* (in press).

[3] GIBSON, W. A. (1956). Nonlinear factors in two dimensions. (Paper read at Amer. Psychol. Ass., Chicago.)

[4] GREEN, B. F. (1951). A general solution for the latent class model of latent structure analysis. *Psychometrika*, XVI, 151–166.

[5] Lazarsfeld, P. F. (1951). The use of mathematical models in the measurement of attitudes. *Research Memorandum RM*–455, RAND Corp.

[6] Patterson, D. G. and Rundquist, E. A. (1933). The occupational background of feeblemindedness. *Amer. J. Psychol.*, XLV, 45, 118–124.

[7] Sabagh, G., Dingman, H. F., Tarjan, G. and Wright, S. W. (1961). Social class and ethnic status of patients admitted to a state hospital for the retarded. *Pac. sociol. Rev.* (in press).

[8] Sarason, S. B. (1953). *Psychological Problems in Mental Deficiency.* (2nd ed.) New York: Harper and Bros.

[9] Spitz, H. H. (1958). The present status of the Kohler–Wallach theory of satiation. *Psychol. Bull.*, LV, 1–28.

[10] Strauss, A. A., and Lehtinen, L. E. (1947). *Psychopathology and Education of the Brain-injured Child.* New York: Greens and Stratton.

[11] Tarjan, G., Wright, S. W., Dingman, H. F. and Sabagh, G. (1961). The natural history of mental deficiency in a state hospital. II. Mentally deficient children admitted to a state hospital prior to their sixth birthday. *J. dis. Child.* (in press).

[12] Thurstone, L. L. (1947). *Multiple Factor Analysis.* Chicago: University of Chicago Press.

Vol. XIV
Part 1
The British Journal of Statistical Psychology
May
1961

THE ROTATION OF PRINCIPAL COMPONENTS

By G. E. R. Burroughs and H. W. L. Miller
University of Birmingham
Shenstone Training College

Doubt has been expressed as to the validity and desirability of rotating principal components. In the present case a battery of tests, which was expected to give clear-cut factors, was analysed by the method of principal components, and the components were rotated orthogonally. It is shown that the results thus obtained were in keeping with those that were expected. It is concluded that, whatever may be the weaknesses of rotation as a technique, it is one which can be applied to principal components quite as well as to factors obtained by other procedures.

I. The Problem of Rotation

With the increasing availability of electronic computors the method of factor analysis known as principal components[1] has now become a feasible undertaking. Already a number of researches have appeared in which this technique has been used with sizable batteries of tests. There seems to be a reluctance, however, to rotate the loadings obtained, and in some quarters a doubt as to whether with this procedure rotation is legitimate. The purpose of this paper is to present a set of data experimentally obtained where it was possible to postulate in advance what factors were to be expected, and to show how far the same factors were arrived at by a principal components analysis followed by an orthogonal rotation.

An analysis into principal components (or, as it is sometimes termed, 'weighted summation') can, like a centroid analysis (the method of 'simple summation'), be carried out with either unities or communalities in the diagonal cells. Which device is adopted depends on whether the factorist is chiefly concerned to relate factors and factor scores to the original test scores, or whether he wishes to describe the battery of inter-correlations in terms of the smallest number of common factors. In a centroid analysis the general practice is to put communalities (or their estimates) into the leading diagonal, though this is not an indispensable feature of the method. With a battery much larger than our own the choice would make little difference to the outcome ; certainly it cannot in itself provide a criterion for the application or non-application of subsequent rotational procedures. In an analysis by principal components it is more usual to take unities as the diagonal entries ; indeed we must do if we want the factor scores to be calculated so that the test scores can be reconstructed from them *exactly*; but again this practice is not an essential feature of the method. Evidently then the decision whether rotation is or is not permissible cannot depend on grounds such as these. What are the reasons for or against using unities in various types of investigation has been discussed at length in an earlier issue of this *Journal* [1, 9]. Since our aim was to offer a demonstration that would be generally acceptable, we ourselves decided to keep to the usual practice.

---

[1] It is unnecessary to argue here whether or not the results of an analysis by principal components may legitimately be termed 'factors'. The origin of the terminology and the problems so raised have been fully discussed elsewhere by Burt (1940), Kendall (1957), Eysenck (1953), and others. In what follows the terms 'component' and 'factor' will be used interchangeably.

The special advantage of using principal components is that each factor accounts for the greatest amount of the total test variance remaining before it is extracted. If unities are used for the self-correlations, then the number of factors required will be exactly equal to the number of tests in the battery (not greater as with the centroid method where specifics are also postulated). These features give the method a strong appeal to the mathematician. Indeed, most mathematicians go further, and apparently hold that, since the maximum amount of the residual variance is accounted for at each stage of factorization, it would be pointless to rotate to positions where this feature was lost. An analogous argument used to be put forward in the case of the centroid procedure: it was said that, since each axis passes through a particular point—a centroid—it would be a pity to lose this feature by subsequent rotation, and that the bipolar factors which it yields give an objective and dichotomous classification based directly on the data which the subsequent rotations are apt to obscure. However, in all such matter, as has been emphasized by Burt [3], one ought to be guided, not merely by mathematical, but also by psychological considerations. Psychologically neither the centroid nor the maximum variance features have any special virtue; they may be considered to be in much the same position as the various transformations which can be used before analysis of variance techniques are applied; that is to say, they are interim devices which facilitate certain subsequent mathematical manipulations.

In an instructive discussion on factor analytic methods [6] Maxwell mentions one feature in particular as contra-indicative of the rotation of principal components, namely, the poor fit which principal components give to the original correlation matrix when the number of factors extracted is fewer than the number of tests. And indeed his objections are frequently quoted by critics when one proposes to supplement an analysis into principal components by an orthogonal rotation to group factors or the like. Now, as we have already stated, with p tests p components are required to reproduce the original correlations exactly; and we would agree that in many cases, if not all, each of the p components ought, if possible, to be computed. But we would suggest that the inability of a limited number of principal components (say the first k) to reproduce the observed correlations exactly is a defect resulting from the limitation in the procedure, not a special weakness to be associated with the subsequent rotations. After all with principal components we have a valid test of significance, due to Bartlett [2], which will guide us in considering when to stop extracting further factors. Often the data obtained are not sufficient to render *all* the factors statistically significant. In such a case a limited number of principal components may give almost as much information as all of them, in regard both to the correlations and the factor scores of the individuals tested (see Thomson [8], p. 117). We think therefore, that a limited number of principal components may be useful, but that this issue has no particular bearing on the rotation problem.

In our opinion the decision whether or not to rotate must turn mainly on one's view of the rotational procedure in itself and not on whether it has been preceded by a principal component, a centroid, or any other well-established method. If one believes that rotation allows of a better or easier psychological interpretation, then, we submit, it can be legitimately used no matter what method of factor analysis has preceded it. The arguments in favour of rotation are not mathematical; and in each research the investigator has to decide its merits on non-mathematical grounds.

II. AN ILLUSTRATIVE EXPERIMENT

Our experiment [7] was concerned with measuring the attitudes of 200 men and women training college students towards the people of four other national or racial groups, viz., American (defined as the white inhabitants of U.S.A.), Chinese, Negroes, and Russians. For each of these the respondents were asked to check a number of items or questions—the same for every group. The questions were presented in three forms—as a Bogardus social distance test, as a Thurstone and Chave attitude test, and as a Likert attitude test. There were thus 12 tests in all.

From the intercorrelations of the tests seven principal components were calculated[1], unities being used as diagonal entries. The correlation table and final residuals are shown in Table I, and the components and their contribution to the variance in Table II. At the time when the analysis was undertaken there was a limit to the number of components which could be taken out in a reasonable period by the machine available ; the decision to stop at seven factors was due partly to this and partly to the small amount of variance remaining to be accounted for at this stage. Bartlett's test showed that all seven components were fully significant, and the trend of the values for chi-squared calculated after each factor had been extracted suggested that one or two significant factors might still remain.

Orthogonal rotations to secure a positive manifold and a maximum number of zero or near-zero loadings were carried out by means of a two-by-two rotation. A modification of Zimmerman's method [10] was used—the new loadings being calculated arithmetically instead of being merely measured on the graph. Rotation to oblique positions might have resulted in a slightly better fit to the

---

[1] We are indebted to Mr. W. L. B. Nixon of the University of London Computor Unit for help in programming and carrying out this task.

Table I. Observed Correlations and Residuals

Test	1	2	3	4	5	6	7	8	9	10	11	12
1		0·318	0·283	0·159	0·119	0·071	0·088	0·034	0·423	0·206	0·229	0·172
2	−0·032		0·777	0·625	0·177	0·468	0·379	0·433	0·297	0·722	0·633	0·565
3	−0·018	−0·057		0·530	0·056	0·367	0·388	0·315	0·250	0·631	0·733	0·494
4	−0·019	−0·011	−0·038		−0·006	0·224	0·114	0·521	0·172	0·458	0·407	0·731
5	0·029	0·035	0·005	−0·017		0·307	0·378	0·229	0·317	0·194	0·156	0·114
6	−0·011	−0·034	0·021	0·055	−0·002		0·572	0·528	0·196	0·557	0·416	0·409
7	−0·013	0·032	−0·045	0·039	−0·012	0·006		0·339	0·134	0·372	0·400	0·201
8	−0·002	0·018	0·048	−0·081	−0·002	−0·060	−0·046		0·096	0·390	0·337	0·625
9	−0·059	0·050	0·024	0·043	−0·038	0·022	0·021	0·010		0·498	0·595	0·330
10	0·028	−0·005	−0·053	0·004	0·008	−0·041	0·019	−0·002	−0·051		0·842	0·648
11	0·035	−0·039	0·004	−0·035	0·027	−0·009	−0·042	0·039	0·022	−0·008		0·612
12	0·025	−0·020	−0·004	−0·038	0·023	0·013	0·006	−0·076	−0·057	0·000	−0·002	

TABLE II. PRINCIPAL COMPONENTS

Type of Test	National Group	I	II	III	IV	V	VI	VII	h^2
Bogardus	1. American	0·334	−0·004	0·657	0·176	0·590	−0·215	0·012	0·969
	2. Chinese	0·844	−0·177	0·022	−0·188	0·194	0·118	−0·247	0·892
	3. Negro	0·785	−0·200	0·085	−0·383	0·150	0·227	−0·025	0·884
	4. Russian	0·668	−0·499	−0·199	0·277	0·135	0·212	0·045	0·877
Thurstone and Chave . .	5. American	0·289	0·725	0·060	0·373	0·000	0·453	−0·171	0·986
	6. Chinese	0·646	0·424	−0·326	−0·082	0·023	−0·399	−0·268	0·942
	7. Negro	0·527	0·578	−0·218	−0·317	0·220	0·039	0·394	0·965
	8. Russian	0·622	0·027	−0·518	0·396	0·110	−0·152	0·092	0·857
Likert	9. American	0·496	0·188	0·641	0·281	−0·302	−0·108	0·113	0·887
	10. Chinese	0·868	−0·003	0·114	−0·142	−0·303	−0·087	−0·135	0·904
	11. Negro	0·836	−0·034	0·201	−0·235	−0·295	0·025	0·139	0·903
	12. Russian	0·787	−0·286	−0·161	0·334	−0·139	−0·050	0·127	0·876
Contribution to Variance (per cent)		44·86	12·31	11·6	8·0	6·5	4·7	3·3	91·2

expected results (some of the negative loadings would probably have disappeared); but orthogonality was retained on account of its simplicity.

III. THE FACTORS EXPECTED AND THE FACTORS OBTAINED

The nature and number of the tests and the nationalities led us to expect factors of four main types. Owing to the limited number actually extracted not all had room to appear. But we postulated that we might get the following.

(i) A general factor indicating the respondent's general attitude to all foreign nationals no matter what type of test was being used (since rotation of the kind we employed usually eliminates the general factor we regarded this merely as a possible but not as a probable factor);

(ii) Four group factors corresponding to the attitudes towards each of the four national groups;

(iii) Three group factors corresponding to each of the three types of test used (we expected greater similarity between the results of different tests than between the results for different nationalities);

(iv) Group factors linking together (*a*) white groups and (*b*) non-white groups respectively. These again we regarded as possible factors only.

The foregoing scheme of factors is shown diagrammatically in Table III.

TABLE III. EXPECTED DISTRIBUTION OF GROUP FACTORS

Test	Factors									
	I	II	III	IV	V	VI	VII	VIII	IX	X
A.1	+	+				+			+	
C.2	+		+			+				+
N.3	+			+		+				+
R.4	+				+	+			+	
A.5	+	+					+		+	
C.6	+		+				+			+
N.7	+			+			+			+
R.8	+				+		+		+	
A.9	+	+						+	+	
C.10	+		+					+		+
N.11	+			+				+		+
R.12	+				+			+	+	

The factors actually obtained are set out in Table IV. So far as possible, they are set out in the same order as those in Table III. The heaviest loadings are printed in semi-bold type. In some cases, however, there are other large factor loadings which tend to detract from the simplicity of the picture; but they are always smaller than the semi-bold figures which appear *in the same column.*

It will be seen that

(i) there is no general factor;

TABLE IV. ROTATED COMPONENTS

Test	I	II	III	IV	V	VI	VII	h²
A.1	**0·635**	0·276	0·219	−0·073	**0·499**	−0·137	−0·408	0·967
C.2	0·055	**0·526**	0·293	0·455	**0·481**	0·237	0·171	0·889
N.3	−0·023	0·418	**0·540**	0·341	**0·452**	0·156	0·267	0·883
R.4	−0·051	−0·011	0·155	**0·785**	**0·480**	0·056	0·011	0·877
A.5	**0·614**	−0·169	−0·210	−0·013	0·111	**0·590**	0·422	0·988
C.6	0·071	**0·540**	0·028	0·236	−0·116	**0·745**	−0·142	0·942
N.7	0·066	0·005	**0·583**	−0·112	0·068	**0·779**	0·014	0·968
R.8	−0·003	−0·047	0·003	**0·647**	0·116	**0·581**	−0·290	0·856
A.9	**0·800**	0·167	0·294	0·290	−0·197	−0·084	0·070	0·889
C.10	0·223	**0·545**	0·374	0·552	−0·061	0·231	0·231	0·902
N.11	0·218	0·374	**0·624**	0·479	−0·043	0·150	0·267	0·902
R.12	0·125	0·069	0·228	**0·868**	−0·124	0·185	−0·046	0·878
Contribution to Variance (per cent)	12·9	11·0	12·5	23·4	8·5	17·2	5·7	91·2

(ii) the first four rotated factors correspond quite well with the four factors postulated above, indicating attitudes towards the four different national groups ;

(iii) of the three group factors postulated to correspond with the types of test employed, two appear quite clearly, namely, factors V and VI, but the group factor corresponding to the Likert tests does not appear (there is in V, however, a sign pattern which suggests that the Likert tests might separate out if more factors had been available for rotation) ;

(iv) the last factor, factor VII, is apparently meaningless; but it is by far the smallest in the seven ; there is, however, a suggestion in the figures that they may serve to distinguish between the American Bogardus test (A.1) and the American Thurstone-and-Chave test (A.5) ;

(v) the remaining factors postulated in Table II, namely factors IX and X, have not appeared.

IV. CONCLUSION

It is not our purpose to discuss in any detail the results obtained from the point of view either of the experimental procedure or of the attitudes revealed. Here our concern has been focused solely on a single point of methodology ; and our conclusion is that the pattern of loadings finally reached is sufficiently close to that which was postulated at the outset to demonstrate beyond question that the rotation of principal components is quite as permissible and reasonable a procedure as the rotation of factors derived by any other procedure.

REFERENCES

[1] BANKS, C. and BURT, C. (1954). The reduced correlation matrix. *Brit. J. Statist. Psychol.*, VII, 107–118.

[2] BARTLETT, M. S. (1950). Tests of significance in factor analysis. *Brit. J. Statist. Psychol.*, III, 77–85.

[3] BURT, C. (1940). *Factors of the Mind.* London : University of London Press.

[4] EYSENCK, H. J. (1953). The logical basis of factor analysis. *Amer. Psychologist*, VIII, 105–113.

[5] KENDALL, M. G. (1957). *A Course in Multivariate Analysis.* London : Griffin.

[6] MAXWELL, A. E. (1957). *A Comparative Study of Some Factor Analytic Techniques.* Ph.D. Thesis, University of Edinburgh.

[7] MILLER, H. W. L. (1960). *Attitudes of Students towards Foreign Nationals.* M.A. Thesis, University of Birmingham.

[8] THOMSON, G. H. (1951). *The Factorial Analysis of Human Ability.* London : University of London Press.

[9] WARBURTON, F. W. (1954). The full factor analysis. *Brit. J. Statist. Psychol.*, VII, 101–106.

[10] ZIMMERMAN, W. S. (1946). A simple graphical method for orthogonal rotation of axes. *Psychometrika*, II, 51–55.

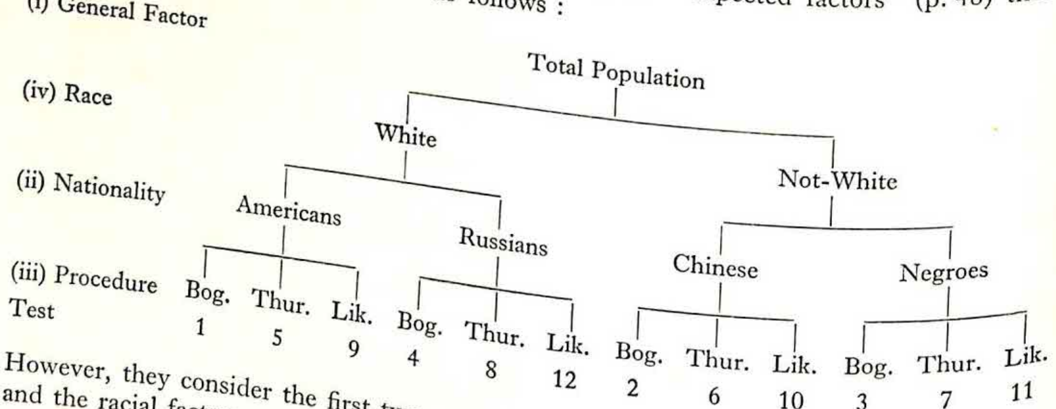
APPENDIX[1]

A Note on Factor Patterns

By CYRIL BURT

The Block-and-Line Pattern. The illustrative example, described in Dr. Burroughs and Mr. Miller's paper, is of special interest to the factorist. Factors are essentially principles of classification ; and the authors' full classificatory scheme furnishes an admirable instance of what may be called *subdivided factors* with *cross classification.* Comparatively few writers recognize how frequently 'subdivided factors' occur in mental testing; and, even when they recognize them, they usually assume the classification is a 'divergent' classification, never wholly or partly a 'cross-classification'.[2]

The divisions and subdivisions postulated by the authors involve four *fundamenta divisions*, as the logicians would say—four 'principles of classification', entailing *one* general factor, *two* racial factors, *three* procedural factors, and *four* nationality factors—ten factors in all to be calculated from a battery with twelve tests (Table III. As formulated by the authors in their preliminary specification of the 'expected factors' (p. 48) their complete scheme may be tabulated as follows :



However, they consider the first two sets of factors mentioned above—the general factor and the racial factors—not as probable but only as possible factors ; and these factors do not appear in their rotated factor matrix (Table IV). We are thus left with a simple factor-scheme of great importance which occurs again and again in mental testing, namely, one in which each of the tests can first be classified according to one particular characteristic—most frequently its *content*—and then cross-classified according to a second characteristic, usually the procedure or *form*. As a rule, the effect of this is at once obvious in the correlation table : one set of factors (here the factors distinguishing the tests according to their procedure) produces blocks of high correlations astride the leading diagonal, while the other set produces one or more oblique rows of high correlations parallel with the diagonal (see Table I above). Among the factor studies previously published in this *Journal* perhaps the clearest instance of this scheme is to be found in A. F. Mahmoud's investigation (VIII, 1955, pp. 121f., Tables II and VI): it relates to eight tests only instead of twelve; but, as we shall see, his correlation table exhibits the same 'block-and-line' scheme as Table I.

---

[1] As will be seen, the following note arose out of certain criticisms made by the referee and quoted on p. 44 below. I am grateful to Dr. Burroughs for his kindness in reading through an earlier draft, and elucidating one or two points about which I was in doubt.

[2] For the terminology, which is borrowed from the logicians' discussion of the theory of classification, see this *Journal*, II, 1949, p. 46.

Possible Criticisms. The immediate purpose of Dr. Burroughs and his colleague, however, was to demonstrate that the factors obtained by weighted summation (' principal components ' as they are sometimes called) can be validly rotated so as to obtain group factors in place of the bipolar factors, much as the factors obtained by simple summation are often rotated. There can, I think, be little question that the results they have obtained fully refute those writers who maintain the more extreme view that such a rotation is in no circumstances justifiable. The authors quote Dr. Maxwell, who mentions " one feature in particular as contra-indicative of the rotation of principal components, namely, the poor fit which principal components give to the original correlation matrix when the number of factors extracted is fewer than the number of tests ". There seems to be a good deal of confusion here. Table I (p. 38) contains 66 observed coefficients obtained by correlating 12 tests. Now with only eight factors we can obtain a *perfect* fit with either principal components (weighted summation) or the centroid method (simple summation), provided suitable communalities are inserted. With only seven factors (as in Table IV above) the fit would not be so good, but it would certainly be better with principal components than with any other method. But possibly, when Dr. Maxwell speaks of ' principal components ', he tacitly assumes that unities not communalities are inserted in the leading diagonal. In that case the fit for 8 factors will undoubtedly be poorer than with communalities. But it will be still decidedly closer than with simple summation, since the method of principal components minimizes the residuals.[1]

However, as has already been pointed out by one critic in correspondence, the analysis carried out by our two authors still leaves a loophole for the more moderate opponents of such rotations. "Dr. Burroughs and Mr. Miller", he writes, "were already aware of the nature and possible resemblances of the several tests and the most likely way in which they would be grouped ; and, as has been frequently shown, when one is armed with this knowledge, it becomes all too easy to rotate the bipolar matrix so as to secure a plausible confirmation of almost any hypothesis one prefers. The task is particularly simple if one is content to end up with one or two meaningless factors (like factor VII in Table IV), and with one or two negative loadings (like those for test A1, test A5, and test R8) to eke out any deficiencies in the fit ; and, when there is a large general factor in the unrotated matrix and none in the hypothetical, there should be no difficulty whatever : for then one can use the general factor loadings to fill out almost any group factors one wishes, and the superfluous overlap can be transferred to other factors. In the present case the numerous positive loadings left after rotation in factors III and IV (which are almost general factors in themselves) suggest that something of this sort has happened ; and as a further result at least one rotated factor and a good many factor loadings appear which were not expected or accounted for in the hypothesis proposed. There is too a marked overlapping in the case of the last four tests, the four loadings in each of the seven factors being (with two minor exceptions) of the same sign in every case : all four tests are of the Likert type, so possibly this indicates a 'Likert factor' which, as the investigators observe, has not been analysed out".

---

[1] In an early issue of this *Journal* (II, 1949, pp. 90f., 'The Alternative Methods of Factor Analysis and their Relations to Pearson's Method of Principal Components') we reproduced the original correlation matrix for which Pearson proposed the method of principal components. A complete factor analysis was carried out by that method, first with unities in the leading diagonal (yielding seven factors for seven ' tests '), then with estimated communalities (yielding four genuine factors, with a negligible fifth to account for residuals). Figures obtained with simple summation are tabulated for comparison ; and the reader can see that the fit is much better with principal components. The results of an orthogonal rotation are also given.

Whether unities or communalities are inserted for the preliminary analysis makes no difference if the analysis is complete 'or virtually complete' so that the discrepancies are non-significant. In either case the same group factor matrix can be obtained by rotation, for, when unities are used, we have only to add a diagonal set of specific factors to bring the communalities up to unity : moreover, the rotation is still orthogonal. Thus, since the initial factor matrix is to be rotated in any case, it makes little difference how it is obtained, as the same group factor matrix can be reached from either type of bipolar matrix. If an electronic computor is available then it is simpler to work with principal components, since nearly all such machines are programmed to carry out this type of matrix-transformation ; if not, simple summation (the ' centroid ' method) will yield the same rotated factors, and at the same time be far quicker for those who only use a desk machine.

The Need for Blind Rotation. I should not myself agree with all these criticisms. To begin with, I think, the investigators have made the aim of confirmation somewhat harder by deciding to dispense with a general factor in the rotated matrix : had they retained it the demonstration would have seemed far more convincing, and many of the meaningless positive loadings in the group factors would have disappeared or been greatly diminished. Secondly, the essential pattern is still obscured by the presence of specific factors which were introduced with the unities and have not been removed by rotation. The fact that most of the observed correlations are well below unity makes it certain that each test must contain a fairly large specific factor ; and by putting unities in the leading diagonal the investigators are forced to convert these to terms of bipolar factors. To secure a distinctive sign pattern for each of the 12 tests help from at least four bipolar factors is needed ; and the modified loadings this produces may well account for the fact that some of the factors and factor loadings seem almost meaningless.

However, the ideal way of meeting the foregoing criticisms would be to ask an independent factorist to carry out a blind rotation. He would then be compelled to base his rotation on the empirical data in front of him, not on his knowledge of the tests or of the particular hypothesis he wished to prove.

The Bipolar Factor Matrix as a Guide. In attempting such a rotation the factorist would, I imagine, find himself obliged to follow the same principle as that adopted by Mahmoud in the article already cited; he would determine the 'grouping' of the 'group-factors' by examining the sign pattern of the bipolar matrix (cf. *loc. cit. sup.*, p. 134 ; the full procedure is described in this *Journal*, III, pp. 53f.). Now, with one or two important exceptions, the whole of the authors' postulated factors can be inferred—so far as the structural pattern is concerned—from the set of unrotated factors shown in Table II.

But, before we attempt to interpret this table, let us recall that, as a general rule, each bipolar factor will classify n tests into two groups with $n/2$ tests in each ; when the rotated factor matrix contains groups with smaller numbers in them, then in the unrotated factor matrix at least two bipolar factors will be needed to express them. Let us now begin with the largest bipolar in Table II (factor II). From what has just been said it must clearly be interpreted in relation to factor V. Together they subdivide the battery into three groups of four tests each. These, as we can see (though our blind factorist cannot), correspond with the three main types of test *procedure* : (factor V contrasts the Likert tests with the other eight, and factor II then subdivides the latter into the Bogardus type and the Thurstone-and-Chave type).

The next largest bipolar is factor III. This must evidently be interpreted in relation to factor IV. Tests 1, 5 and 9 have positive loadings in both. Tests 4, 8 and 12 have positive loadings in factor IV, but negative loadings in factor III. This gives two further groups, each containing three tests. (Knowing the content of the tests, we can identify those just mentioned as tests relating to the Americans and to the Russians respectively.) Factors VI and VII distinguish between the Chinese and the Negroes, except in the Bogardus group, where there seems to be no evidence for such a distinction: (in this case I find it difficult to imagine any factorist introducing such a distinction as indicated by the loadings in bold type in Table IV, unless he already knew that his hypothesis required it; this is surely a point where a blind factorist would *not* verify the expected scheme).

Dr. Burroughs and Mr. Miller apparently found no indication of the group factors linking together (*a*) white groups and (*b*) non-white groups (factors IX and X). But factor IV in the bipolar matrix appears to do precisely this.

They are also puzzled, so they tell us, by the nature of their last and smallest factor—factor VII. In this only two loadings are of any appreciable size—one negative (−0·408 for test A1) and one positive (+0·422 for test A5). This produces a negative correlation between tests A1 and A5. Now the first factor of all—factor I—has exceptionally large loadings for the three 'American' tests—A1, A5, and A9. This is in keeping with the table of *expected* factors. But there is nothing either in the *observed* correlations or in the unrotated factors to support so large a group factor. The correlation between A1 and A5 is only 0·119 ; and the other two correlations are not much larger. Clearly factor VII with its big negative loading is required merely to cancel the effect of the two large positive loadings assumed for factor I. The negative loadings for A5 in factor II and for A9 in factor V have a similar purpose. Had the three loadings in factor I been smaller there would have been no need to introduce these negative loadings.

The Correlation Table as a Guide. The expected distribution of the group factors shown in Table II happens to be of a type which, when present, commonly shows up fairly distinctly in the matrix of observed correlations. Let us look once again at the correlation table and accompanying factor matrix printed in Mahmoud's paper. Except that there are only 2 × 4 tests instead of 3 × 4, the factor pattern (*loc. cit.*, p. 133, Table VI) is almost identical with that shown in Table III above. The 'basic factor' in the former corresponds with factor I in the latter; the next two factors (group factors with four loadings, representing different types of test content) correspond with factors VI and VII ; the last four ("quasi-specifics peculiar to different *forms* of test", with loadings arranged obliquely) correspond, as the scheme is now arranged, with factors II, III, IV, and V, and the two remaining factors (for the 1st and 2nd days) correspond with factors IX and X. Naturally there is nothing answering to factor VIII, since this represents a *third* application of the four subtests redesigned to conform with a third type of test procedure (in the present case, the Likert).

When we turn to the two correlation matrices, the correspondence is almost as striking. In both tables the pattern of high coefficients produced by the pattern of factor loadings stands out quite clearly. Indeed it is, if anything, still more conspicuous in Table I above (p. 38) because of its larger size. First, factors II, III, IV, and V (p. 40) leave two oblique ridges of high correlations, parallel to the high self-correlations ordinarily inserted in the leading diagonal, and running from the correlation between tests 1 and 5 down to that between tests 8 and 12 and again from the correlations between tests 1 and 9 to that between tests 4 and 12. Secondly, factors VI, VII, and VIII each produce a 4 × 4 square of high correlations astride the leading diagonal. In addition to this factor X is responsible for the large isolated correlations between the tests relating to the Chinese and the Negroes (tests 2 and 3, 6 and 7, 10 and 11, etc.). Factor IX, according to the hypothesis proposed, ought to have the same effect on the tests relating to the Americans and Russians ; but it does not.

Discrepancies between Data and Hypothesis. The popularity of the 'hypothetico-deductive' theory of scientific proof has led many research students to suppose that the mark of a successful experiment is to find empirical data agreeing in detail with the hypothesis which the investigator wishes to establish. But the really instructive features of an inquiry are more likely to be those unexpected discoveries which were not contemplated in, or deducible from, the initial hypotheses. It is from the discrepancies which seem to correct or contradict our preconceived notions that we learn most.

In the present investigation the expected classification appears to be contradicted at two significant points. First, we find that, when the two 'white nations' are American and Russian, the expected group 'factor linking white groups' does not appear. Secondly, we note that the correlation between the attitudes towards the Chinese and the Russians is almost as high as the correlation between the attitudes towards the Chinese and the Negroes. What inferences are to be drawn from these two discrepancies can hardly be decided without reference to the actual questions that were asked. It may be that some popular antithesis, like that between 'communist' countries and 'non-communist' or between groups with a 'Western' way of life and a 'non-Western' way has been more influential in determining the students' attitude than any simple distinction according to colour or race. But that is a problem which could hardly be solved by further statistical analysis of the data that have been given.

Computing Methods. Whenever the battery of tests is large and the factor pattern therefore somewhat intricate, most factorists, it would seem, find it difficult to trace the clues to the arrangement of the group matrix in the initial bipolar matrix. It may therefore be helpful to examine an illustrative instance of the pattern with which we have here been concerned, reduced to its simplest possible terms. For purposes of demonstration it is usually best to take *exact* figures for the hypothetical scheme and to assume that the loadings in the same factor are always equal. The relations then stand out unobscured, and the accuracy of the rotation can be judged beyond question.

1. *Bipolar Analysis.* Table I A gives the smallest possible correlation table, which will show the same type of pattern as that displayed in Table I on p. 38. The 'blocks' of augmented coefficients are printed in semi-bold, and the oblique 'lines' in italic. A complete analysis into principal components has been carried out, and the resulting factor loadings are shown in Table II A. The sign pattern in the table can be compared with that set out in Table II above.

The scheme of classification which the factor pattern implies is here unmistakable. We have (i) a general or 'basic' factor common to all four tests; (ii) two fairly large group factors common to (*a*) tests 1 and 2, and (*b*) tests 3 and 4; (iii) two slightly smaller group factors common to (*a*) tests 1 and 3, and (*b*) tests 2 and 4; (iv) finally the fact that there is a unique pattern of signs for each of the four tests indicates that each has its own specific. All this we might have inferred by inspecting the correlation table, though with less confidence.

2. *Group Factor Analysis.* To determine the matrix of group factors we begin by picking out correlations between tests which, according to our scheme, could be due only to the single general factor, e.g., the correlation between tests 1 and 4. With no more than one pair of tests we must assume equal factor loadings, viz., $\sqrt{(0{\cdot}500)}=0{\cdot}707$. For the other factors we have $\sqrt{(0{\cdot}750-0{\cdot}707^2)}=0{\cdot}500$, $\sqrt{(0{\cdot}660-0{\cdot}707^2)}=0{\cdot}400$, and $\sqrt{(1{\cdot}000-0{\cdot}707^2-0{\cdot}500^2-0{\cdot}400^2)}=0{\cdot}300$. The complete results are shown in Table IV A.

We could proceed in exactly the same way with Table I in the preceding paper. It is plain, for instance, that, if we accept the hypothetical scheme shown in Table III, then the low correlations of test 1 with tests 6, 7, 10, and 11 must be due to the general factor only. With this large battery we have for each test as many as *four* correlations due solely to this one factor; we must therefore proceed by averaging (see this *Journal*, III, pp. 47f., for details). After subtracting the effect of the general factor we next calculate the loadings for factors IX and X. For this we have two correlations for each test (e.g., those of test 1 with 8 and 12). Next we can calculate the loadings for the three main blocks (factors VI, VII, and VIII) and finally those for factors II, III, IV, and V.

3. *Rotation.* Returning to our artificial example we must now attempt to reconstruct the original correlations from the new group factor matrix, and calculate the various residuals. If (as will probably be the case with an empirical correlation table) some of the residuals are fully significant and thus imply some measure of overlap, we must next proceed to a formal rotation. The requisite rotation matrix is given in Table III A. Calculation will show that the sum of squares for each row is 1·000 and the product-sum for each pair of rows is zero : in other words, the table consists of the first four rows of a 9 × 9 orthogonal matrix. On multiplying the factor matrix for principal components by this rotation matrix we obtain the loadings set out in Table IV A. Our hypothetical scheme is thus confirmed; and, since the correlation table (Table I A) was in fact constructed from Table IV A, the accuracy of the results demonstrates the validity of applying orthogonal rotation to a factor matrix obtained for principal components.

4. *The Reduced Correlation Matrix.* I now want to show that the same result could have been obtained much more speedily with the more usual procedure. We begin by inserting those particular self-correlations (or 'communalities') which will reduce the correlation matrix to its lowest possible rank—here a rank of three[1]. The requisite values are 0·91. We then factorize this reduced matrix by simple summation, and obtain a bipolar matrix containing three factors only instead of four (Table II B). This implies the same classificatory scheme as before ; but the rotation matrix will be different. It now contains only five columns instead of nine (Table III B) ; but the square-sums are still unity and the product-sums zero. On postmultiplying Table II B by this orthogonal matrix we obtain the same loadings for the group factors as before, but the specific factor loadings must now be calculated by the formula $\sqrt{(1-h^2)} = \sqrt{(1-0\cdot91)} = 0\cdot30$.

So far as theoretical considerations go, therefore, it would seem to be almost a matter of indifference whether the rotation starts from a matrix of principal components obtained by weighted summation or from a matrix of centroid factors obtained by simple summation, or again whether we start with unities or with communalities in the leading diagonal. With communalities, the factor pattern is likely to show up rather more clearly in the bipolar matrix. But in the main the choice will, in my view, rest chiefly on practical grounds. For the ordinary research-student equipped only with a desk-machine, factorization by simple summation applied to a table with communalities will prove by far the simplest and quickest procedure, and as accurate as any other method if the results are eventually to be rotated.

---

[1] In theory, so long as we are free to choose the diagonal entries, a 4 × 4 matrix should be reducible to rank two. But on solving the requisite determinantal equations we obtain for the diagonal entries 1·4717 (twice) and 0·5083 (twice). This certainly yields two factors only, viz.

I. 0·9929, 0·7100, 0·7100, 0·9929,
and II. 0·6972, 0·0645, −0·0645, −0·6972.

But for a correlation matrix 1·4717 would be an impossible value.

In the illustrative example here chosen the results of simple and weighted summation are the same. But this does not invalidate the argument. If any reader feels a doubt on this point, then he can easily construct a suitable example for himself. He must take *different* values for the test-loadings in the same general or group factor. This will produce a slightly more complex matrix and entail a slightly more complicated set of calculations, and so obscure the points I wished to bring out ; but would give a clearer answer on this further issue, if it were raised.

TABLE I A. CORRELATIONS

Test	*i*	ii	iii	iv
i	**1·00**	**0·75**	*0·66*	0·50
ii	**0·75**	**1·00**	0·50	*0·66*
iii	*0·66*	0·50	**1·00**	**0·75**
iv	0·50	*0·66*	**0·75**	**1·00**

TABLE I B. CORRELATIONS

Test	***i***	ii	iii	iv
i	**0·91**	**0·75**	*0·66*	0·50
ii	**0·75**	**0·91**	0·50	*0·66*
iii	*0·66*	0·50	**0·91**	**0·75**
iv	0·50	*0·66*	**0·75**	**0·91**

TABLE II A. PRINCIPAL COMPONENTS

Factor	I	II	III	IV
i	+0·8529	+0·3841	+0·3202	+0·1500
ii	+0·8529	+0·3841	−0·3202	−0·1500
iii	+0·8529	−0·3841	+0·3202	−0·1500
iv	+0·8529	−0·3841	−0·3202	+0·1500

TABLE II B. CENTROID FACTORS

Factor	I	II	III
i	+0·8396	+0·3535	+0·2828
ii	+0·8396	+0·3535	−0·2828
iii	+0·8396	−0·3535	+0·2828
iv	+0·8396	−0·3535	−0·2828

TABLE III A. ROTATION MATRIX

Factor	G	Gr 1	Gr 2	Gr 3	Gr 4	Sp 1	Sp 2	Sp 3	Sp 4
I	+0·6873	+0·0859	+0·0859	+0·0550	+0·0550	+0·0077	+0·0077	+0·0077	+0·0077
II		+0·4237	−0·4237			+0·0381	+0·0381	−0·0381	−0·0381
III				+0·3903	−0·3903	+0·0549	−0·0549	+0·0549	−0·0549
IV						+0·2500	−0·2500	−0·2500	+0·2500

TABLE III B. ROTATION MATRIX

Factor	G	Gr 1	Gr 2	Gr 3	Gr 4
I	+0·8420	+0·2977	+0·2977	+0·2320	+0·2320
II		+0·7071	−0·7071		
III				+0·7071	−0·7071

TABLE IV A. GROUP FACTOR MATRIX

Factor	G	Gr 1	Gr 2	Gr 3	Gr 4	Sp 1	Sp 2	Sp 3	Sp 4
i	0·7071	0·5000		0·4000		0·3000			
ii	0·7071	0·5000			0·4000		0·3000		
iii	0·7071		0·5000	0·4000				0·3000	
iv	0·7071		0·5000		0·4000				0·3000

Vol. XIV
Part 1
The British Journal of Statistical Psychology
May
1961

NOTES AND CORRESPONDENCE

FACTOR ANALYSIS AND ITS NEUROLOGICAL BASIS

To the Editor, British Journal of Statistical Psychology

SIR,—In the discussion with Mr. Taylor, Mr. Hansel, and Mr. Clarke, on the methodology appropriate for the statistical psychologist, both Mr. Gregory and Professor Burt have appealed to the so-called ' uncertainty principle '. They believe that this principle operates within the nervous system itself, and so implies the presence of an irreducible random element in all psychophysical reactions.

Now one very familiar technique which is based on the hypothesis of random reactions is Thomson's ' sampling theory ', described in his *Factorial Analysis of Human Ability*, and interpreted in concrete neurological terms. Guilford in his *Psychometric Methods* (pp. 465 *et seq.*) enumerates its many merits, and commends it as a basis both for theory and practice. Still more recently Herbert Solomon has approved it as " framing and formulating a new theory of mental ability " (*Mathematical Thinking in the Measurement of Behaviour*, 1960, pp. 285 *et seq.*). It is therefore surprising that, although the *British Journal of Statistical Psychology* has discussed almost every other method of factor analysis, it scarcely ever mentions this model. Since it claims to publish, not only current research, but also historical studies indicating the problems and procedures from which current researches have arisen and expository notes and papers for the ordinary student explaining the techniques employed, surely the sampling theory deserves a brief description.

Journals of statistical psychology are too apt to restrict themselves to what Guilford has called " exercises in statistical manipulation "[1], and to concentrate on abstract mathematical models and methods for their own sake with little or no reference to the concrete neurological processes which must after all be thought of as determining the statistical variables or factors involved if those factors and variables are to have any psychological meaning. For the contemporary behaviourist the special advantage of the sampling theory is that it offers an explicit physiological framework for statistical techniques and for the construction of psychological tests and psychiatric hypotheses. It links the results of the factorist with the actual activities that presumably go on within the brain. For the contemporary psychiatrist it provides an attractive explanation of mental breakdown far superior to those put forward by other factorists. According to Spearman, if I understand him aright, mental breakdown is the result of excessive tension set up by excessive cerebral energy, resulting in what Burt calls mental ' turbulence ' (though, if we are to have physical analogues or models, I should have thought models drawn from electrical theory more appropriate than hypothetical models or metaphors derived from hydrodynamics). According to Thomson's theory mental breakdown and confusion would be the result of an excessive random element producing chaotic reactions—a far more plausible explanation.

Moreover, factors as such need no longer be reified. In their place we have a sampling of nerve-cells and neural arcs. Tests excite chains of stimulus and response. Abilities are no mere faculties : they are just samples of nerve-cells. And the essential nature of a test is "to activate a random sample of the nerve-cells of the brain". Further, as we are now told, all learning—even the learning of a conditioned reflex—proceeds by trial and error, or, in other words, by sampling—random at first, and progressively restricted by 'reinforcement' during the phase of practice. This is admitted by Burt and Foley in their "Statistical Analysis of the Learning Process" (this *Journal*, IX, pp. 49 *et seq.*); but they make no attempt to interpret their assumptions in neural or cerebral terms. If Professor Burt rejects Thomson's very attractive neurological hypothesis, what alternative would he suggest in its place?

---

[1] See his criticisms of previous articles in *Psychometrika* and other journals of statistical psychology in the current issue of that periodical (XXV, 1961, p.122).

In his discussions on the 'uncertainty principle' Professor Burt seems rather to shirk the problem from which it all stemmed—the problem of 'will as such'. I suggest that what seems like the 'certainty of will' is really an effect of learning, operating to reduce the uncertainties of trial and error. The brain thus becomes the 'organ of will'.

May I add that, so far as I can see, Thomson himself does not assume any *special* type of 'uncertainty' operating in the brain? For him the sampling of 'bonds' or nerve-cells in no way differs from the sampling of cards or dice. As a behaviourist I should hold that *all* the processes of the brain—including even the slips and accidents that every examinee is apt to make—are rigorously determined, just like the fall of a coin or a die. If the physiological psychologist maintains the contrary, he should explain in detail how he conceives the processes that he postulates. And perhaps in his reply he will remember that most statistical psychologists, and even, I would say, most psychiatrists, really know very little about the working of the brain other than what can be found in the elementary textbooks of physiology which they thumbed over in their student days.

The brain, the nerve-currents in the brain, and the material stimuli that excite the brain, are all of them physical entities in a three-dimensional physical space. The sensory percepts of the person who owns the brain also form a spatio-temporal continuum. Is it not therefore possible to regard these various patterned spaces as part of one all-embracing space (or rather 'space-time')? And in that case could not this hyperspace be regarded as a kind of 'factorial space' like that depicted in the three-dimensional diagrams printed by Thomson, Burt, and Cattell—e.g., *Brit. J. Statist. Psychol.*, VIII, p. 110? Here the first dimension represents *time*; the second (a non-personal dimension allocated to tests or stimuli) really represents a projection (on a single line) of *physical space* and the events within it; and the third, labelled persons, represents in linear projection the various *mental spaces* of individuals concerned. But since both physical space and each mental space are three-dimensional, we have a psychophysical factor-space which, like that described by Dr. Smythies, has in all $3(m+1)$ dimensions, m being the number of individuals (cf. his *Analysis of Perception*, 1956, pp. 29, 129f.). The study of this psychophysical factor-space is surely a task for the statistical psychologist, and, if successful, might lay the foundations for a psychophysical cosmology that would take the place of traditional metaphysics.

This mode of interpreting factorial results comes nearer to that adopted by Thomson than that favoured by Spearman or Thurstone or Burt, and again illustrates the fruitfulness of the sampling theory. The theory easily covers the problem of individual differences, since according to Thomson tests may be regarded as 'sampling men' as well as 'sampling abilities', though he himself does not work out this suggestion. However he accepts Burt's conception of 'person factors' in a 'person space' and 'test factors' in a 'test space', as well as his contention that there are certain 'reciprocal relations' uniting the two. It would seem therefore to follow, on Thomson's own principles, that both kinds of space—the individual's space and the space of physical stimuli or tests—become partial subspaces of a common hyperspace.

All this implies three overruling neurological conditions which the statistical psychologist should keep in mind in constructing the abstract theories which he proposes to test, viz., the way the brain functions in perception, in learning, and in what is popularly called willing. To each, if Thomson is right, the clue is random sampling.

LEWIS R. LEVY.

FACTOR ANALYSIS AND ITS NEUROLOGICAL BASIS

A Reply

By Cyril Burt

Dr. Levy has raised a variety of questions all of which undoubtedly deserve fuller consideration from statistical psychologists than they have hitherto received. Nevertheless, many are primarily problems in general psychology or neurology, and can therefore only be touched on very briefly here. In the main I shall confine myself to those questions which have a fairly direct bearing on factorial techniques and on the interpretation of psychological 'factors', and leave the remainder to another occasion.

I. The Sampling Theory

The Sampling of Neural Bonds. Thomson's sampling theory was originally put forward as an alternative to Spearman's one-factor theory. Spearman's model sought to explain the correlations between mental tests without invoking group factors; Thomson's model sought to explain them without invoking any general factor. His own belief, so he tells us, was that the "near-hierarchical arrangement", on which Spearman laid so much stress, was "produced by a number of group-factors which are of wide range and not mutually exclusive" (i.e., which overlap).[1] There was a third and still older theory—the theory proposed by Galton and adopted by McDougall, Flugel, and myself—involving the two-fold assumption of a single general ability supplemented by a number of special aptitudes, "a combination of old-fashioned faculties with a new-fangled super-faculty", as Thomson described it. This, so he contended, was in conflict with the current conception of the mind or brain inculcated by the associationist and behaviourist schools which were then once more in the ascendant.

Spearman himself identified the general factor with 'cerebral energy'. I identified it with the general structure of the brain. Thomson, challenged to interpret his theory in neurological terms, suggested that his "factors or elements" could be "identified on the mental side with Thorndike's associations or 'bonds', and on the bodily side with neurons or neuron-arcs". And he argued that, "so far from showing that the mind or brain has a definite structure" (as McDougall and his followers held[2]), "the low rank of the near-hierarchical correlation tables really means that it has hardly any structure at all".[3] Thomson in short adopts what Sholl has called the 'bowl-of-porridge' conception of the brain.

The nerve-cells or neuron-arcs are described as "firing on an all-or-none basis". Thus, unlike the older type of group-factor, Thomson's 'elements' or 'factors' have no quantitative measurements attached: each counts for 1 if it influences performance at a test and for 0 if it does not. The brain is supposed to possess a large but finite number of these 'elements'; and every test takes a random sample of them. Nevertheless, different tests differ greatly in 'strength or richness': some take large samples, others take small. And in their sampling operations they are very apt to overlap, the 'richest' naturally tending to overlap the most.

Hierarchy without a General Factor. The details of the theory will become clearer if we examine Thomson's discussion of his own example in the light of his earlier papers.

---

[1] W. Brown and G. H. Thomson, *The Essentials of Mental Measurement*, 1925, pp. 174f.

[2] W. McDougall, *Psychology*, 1912, chap. III, 'The Structure of the Mind': cf. C. Burt, 'The Structure of the Mind: A Review of the Results of Factor Analysis', *Brit. J. Educ. Psychol.*, XIX, 1949, pp. 176f., and refs.

[3] G. H. Thomson, *The Factorial Analysis of Human Ability*, 3rd ed., 1948, pp. 303f.

This, I think, will give a view of his theory somewhat different from that which Dr. Levy has apparently derived from brief accounts in American textbooks. For simplicity, Thomson supposes that the mind "can form in all 108 bonds or some multiple of 108 (the number is chosen merely because it leads to no fractions)". Four tests are applied. They sample 5/6, 4/6, 3/6, and 2/6 respectively of the total number of bonds; i.e., the samples will contain 90, 72, 54, and 36 bonds respectively. Now "these fractions", we are told, "are the same in value as the communalities"; and on working out the correlations we obtain the coefficients shown below. To elucidate the way Thomson's table (*op. cit.*, p. 43) has been constructed I have inserted both the communalities and the factor saturations: the latter can be obtained by using either 'weighted summation' (with the vulgar fractions) or 'simple summation' (with the equivalent decimal fractions).[1]

TABLE I. THOMSON'S TABLE OF CORRELATIONS

Communalities and Saturations for the General Factor added by Burt

Test		1	2	3	4
Factor Saturation		$\surd(5/6)$	$\surd(4/6)$	$\surd(3/6)$	$\surd(2/6)$
1	$\surd(5/6)$	[5/6]	$\surd 20/6$	$\surd 15/6$	$\surd 10/6$
2	$\surd(4/6)$	$\surd 20/6$	[4/6]	$\surd 12/6$	$\surd 8/6$
3	$\surd(3/6)$	$\surd 15/6$	$\surd 12/6$	[3/6]	$\surd 6/6$
4	$\surd(2/6)$	$\surd 10/6$	$\surd 8/6$	$\surd 6/6$	[2/6]

"The tetrad-differences", as Thomson points out, "are all zero by inspection"; hence the table *could* be explained by a single general factor. Thomson himself however puts forward an "alternative explanation" which assumes that the size of each correlation is determined by the number of bonds common to the two tests. The amount of overlap he demonstrates by constructing a diagram of overlapping ovals, rather like those popularized by Euler to illustrate deductive arguments: the overlapping parts he calls 'cells' (with no allusion of course to nerve-cells). With the experimental data before us we can calculate the 'richness' of each test (i.e., their communalities); but we cannot "tell *with certainty* what the contents of each cell will be; we can only calculate the *most probable* contents".

We proceed as follows. "The number of bonds used in all four tests is most probably

$$\tfrac{5}{6} \times \tfrac{4}{6} \times \tfrac{3}{6} \times \tfrac{2}{6} \times 108 = 10 \text{ bonds}".$$

Similarly the most probable number used in the first three tests (tests 1, 2, and 3) will be

$$\tfrac{5}{6} \times \tfrac{4}{6} \times \tfrac{3}{6} \times 108 = 30 \text{ bonds}.$$

But 10 of these have already been included in the set common to all four: that leaves only 20 bonds entering into these three and no more. Similarly there will be 10 bonds entering into tests 1, 2, and 4 and no more. Finally, the number entering into 1 and 2 will be

$$\tfrac{5}{6} \times \tfrac{4}{6} \times 108 = 60 \text{ bonds}.$$

But 40 of these have already entered into tests 1, 2 and either 3 or 4 or both. That leaves only 20 peculiar to just these two.

The final result will be clearer if we enter the figures in the columns of a table rather than in a diagram of ovals. They are set out in Table II. The arabic figures (1, 2, 3 or 4) in the top half of the table indicate which tests enter into the group factor specified, and the figures in the bottom half the number of ' bonds ': thus 10 bonds are common to all four tests, 20 to the first three only, and so on. The figures in the last column but one show the number of ' bonds ' specific to each factor; except for reasons of space each should be printed with a column to itself.

---

[1] This method of accounting for correlations appears to have been suggested by experiments carried out by Weldon in Pearson's laboratory. His object was to obtain correlations by throwing known samples of dice or coins and treating a certain number as common to both correlated variables: see W. F. R. Weldon, 'Inheritance in Animals and Plants', ap. *Lectures on the Method of Science* (T. B. Strong, ed.), 1906, pp. 100f. Thomson's earliest attempts at producing a 'hierarchy with no general factor' were carried out by throwing dice to represent 'common' and 'specific' factors exactly as Weldon had done, and using Weldon's formula.

TABLE II. GROUP FACTOR MODEL: NUMBER OF OVERLAPPING BONDS

	Group Factors											Specific Factors	Total
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII to XV	
	1	1	1	1	1	1	1	–	–	–	–		
	2	2	2	–	2	–	–	2	2	2	–		
	3	3	–	3	–	3	–	3	3	–	3		
	4	–	4	4	–	–	4	4	–	4	4		
Test													
1	10	20	10	5	20	10	5	–	–	–	–	10	90
2	10	20	10	–	20	–	–	2	4	2	–	4	72
3	10	20	–	5	–	10	–	2	4	–	1	2	54
4	10	–	10	5	–	–	5	2	–	2	1	1	36

It will be seen that Thomson's method of analysis postulates as many factors as there are combinations of tests taken n, $(n-1)$, $(n-2)$, ... or 1 at a time: i.e., 2^n-1 factors in all; of these n will be specific: some of course might have *no* bonds. Here, for example, he assumes the operation of 11 overlapping group factors and 4 specific factors—15 factors in all. The numbers in the lower half of the table can be regarded as indicating the variances contributed by the several factors: hence $\sqrt{(10/90)}$, $\sqrt{(20/90)}$, . . . , $\sqrt{(10/72)}$, $\sqrt{(20/72)}$, . . . , etc., would represent their factor loadings. Thomson, however, prefers to express each correlation, not as a product of the factor loadings, but rather in terms of the total number of common bonds; and for that purpose he borrows Weldon's formula

$$r_{ij}=\frac{\text{Number of elements common to both tests}}{\text{Geometric mean of total number contained in each test}}$$

$$=\frac{c}{\sqrt{(t_i\times t_j)}}=\frac{c}{\sqrt{\{(c+s_i)(c+s_j)\}}}. \tag{2}$$

Thus from the table above (or from the corresponding diagram in Thomson's chapter) we can count up that there are $10+20+10+20=60$ elements or bonds common to tests 1 and 2. Hence

$$r_{12}=\frac{60}{\sqrt{(90\times 72)}}=\frac{\sqrt{20}}{6}.$$

But evidently, since we really need to know only the *total* number of bonds common to the two tests (60 in the case of tests 1 and 2), there is no need to split it into a series of additive portions as shown in the table or Thomson's diagram.

Alternative Interpretations. Let us now express Thomson's procedure in the more familiar notation. If r_{ig} denotes what would ordinarily be called the general factor loading for the ith test (first column in Table I), then r^2_{ig} will denote the factor variance (or ' communality ', as Thomson prefers to call it); and if T designates the total number of bonds available (here 108), then " the number of bonds required by the test " will be $r^2_{ig}T$; and ' the most probable number of bonds common to two tests ' will be $r^2_{ig}\,r^2_{jg}\,T$.

Thus, by eqn. (2),

$$r_{ij}=\frac{r^2_{ig}\,r^2_{jg}\,T}{\sqrt{\{(r^2_{ig}\,T)(r^2_{jg}\,T)\}}}=r_{ig}\,r_{jg}. \tag{3}$$

But this is the ordinary product-formula for calculating a correlation on the assumption that only a single general factor is involved. Hence Thomson's procedure is here algebraically identical with that commonly adopted for factorizing a correlation-table of rank one into a series of saturations (or loadings) for one common factor only.

Implications of the Theory. Thomson's sampling theory, when it first appeared, did much to bring home to psychologists the real nature of correlation. Up to then the popular notion had been derived from the early Galton–Spearman analysis of a single correlation coefficient with only one common factor and an independent specific factor, one peculiar to each test. Thomson's approach emphasized the need to consider the whole matrix of intercorrelations, and to regard each and all the correlated variables as being, quite possibly, functions of a number of 'causal elements' or 'factors'.[1] The notion that "mental activity depended for its efficiency on the working of a large number of brain-cells" provided a convenient illustrative basis for a concrete exposition of the mathematical derivation—a derivation which few psychologists in those days could follow when stated solely in the abstract notation of the integral calculus and determinantal equations, after the fashion regularly adopted by Karl Pearson. But I find it difficult to believe that as "a neurological interpretation of factor theory" the hypothesis can claim any literal validity today.

The chief difference between Thomson's approach and my own may, I think, be stated as follows. Thomson was interested primarily in the description of the *abilities* tested and in the differences between those abilities; I was interested rather in the *persons* tested and in the differences between them. In Thomson's theory individual differences play no part, though, of course, the theory does not exclude them. He himself, however, always speaks of '*the* mind', as if the successive applications of the tests sampled the same mind on each occasion, not different minds.[2]

---

[1] This was essentially Pearson's approach as contrasted with Galton's: see Brown's chapter on the General Theory of Multiple Correlation ("based on a lecture demonstration given by Professor Pearson") in the first edition of *The Essentials of Mental Measurement* (pp. 128–133: The chapter was unfortunately omitted from later editions as too abstruse and sufficiently covered by the chapters on sampling theory).

I fully agree with Guilford's detailed statement of the superior merits possessed by Thomson's 'sampling theories' as compared with Spearman's 'two factor theory'. On a later page Guilford himself seems ultimately to decide in favour of a 'weighted group-factor theory', and evidently endorses most of the current criticisms of the sampling theory (*op. cit.*, pp. 467f.). It is too often forgotten that Thomson's theory, like Spearman's, was originally put forward to explain why so many tables of test-correlations had a rank of one and no more. My own answer was: "not because of the way the tests sample the 'bonds', but because of the way the psychologists have sampled the tests". In the early small scale studies, the investigator nearly always avoided duplicating his tests, and selected *one* specimen only of each type of cognitive activity. As soon as psychologists started to use large batteries and groups of *similar* tests, then group-factors as well as the single general factor began to appear (see Burt, *Brit. J. Psychol.*, XXX, 1939, pp. 85f.).

[2] This criticism was also urged by Ballard, who, as a school inspector, was equally interested in differences between the pupils tested. Thomson replied by saying that there were no reasons why the theory *must* imply that individual minds were all the same: like the tests, he says, men might also "be samples"; and, like the tests, "some might be rich, some poor in the bonds". But this assumption nowhere enters into the structure of his theory, nor are we told how the 'richness' of each man should be assessed. (Several alternative devices are conceivable, analogous to those used in ordinary factor theory; but we do not know which of them, if any, Thomson would have used.)

If we include the possibility that individuals differ in the total number of bonds they possess (or, as I would rather say, the number they can form), then Thomson's account of the general factor does not vary greatly from my own. In my view the predominance of the general factor simply implies that a given person's brain or nervous system is more or less of the same quality throughout: and this is very much what Thomson endeavoured to convey by declaring that the mind or brain was 'structureless'.[1]

II. The Neurological Basis

Neural Structure and Factorial Structure. Dr. Levy asks, "if Burt rejects Thomson's very attractive neurological hypothesis, what alternative would he suggest in its place?" My own view was briefly outlined in one or two earlier papers, and was based on the conception of the central nervous system and its working elaborated by Sherrington. Later work enables us to fill in fuller details.[2]

The central nervous system may be pictured as an immense tridimensional network, consisting of elongated cells (or 'neurons') and their branches, joined end to end by cell-junctions (or, to use Sherrington's term, 'synapses'). The neuron is a cell specifically adapted for propagating electro-chemical impulses. Hence the whole system forms a highly complicated set of interlacing paths or channels. It is not entirely 'undifferentiated', and by no means completely 'structureless'. Within the overgrown nerve-ganglion that we call the brain some of the nerve-cells are collected together in certain central cell masses, such as the optic thalamus and the corpus striatum, but most of them are located in the crumpled outer sheet of grey matter, known as the 'cortex', where they can make easy contact with each other. Consciousness seems to be associated solely with what goes on in these particular regions.

The areas of the cortex may be classified into two main groups—areas directly connected by nerve-fibres to remoter parts of the nervous system and so ultimately with the sense-organs and muscles (the so-called 'projection areas'), and areas *without* such connections (the so-called 'association areas'). The whole structure is arranged in a hierarchy of 'levels', and reveals slight but unmistakable differences in the different areas. The levels appear to correspond with the group factors for mental *processes* (sensori-motor, associative, relational, and the like); and the areas to group factors for different types of mental *content* (visual, auditory, kinaesthetic, verbal, etc.). But both in the higher mammals and in man the general differences between different individual brains are far more striking than the more specific differences between different parts belonging to the same brain. Thus, as compared with that of the normal person, the cortical tissue of the low grade defective displays comparatively few cells, and these have comparatively few branches; and both cells and branches appear to be irregularly arranged. These distinctive peculiarities appear in all areas and in all levels of the same individual brain. Such a brain is therefore lacking what may be called systematic complexity. It is this general character of the individual's brain-tissue—viz., the general degree of systematic complexity in the neuronal architecture—that seems to me to represent the general factor, and account for the high positive correlations obtained between various cognitive tests.[1]

---

[1] I recognized minor local variations in different regions. Thomson in the later versions of his theory virtually admitted this contention by suggesting the possibility of 'sub-pools', though he is inclined to believe that these are mainly due to "environment and education and life".

[2] Cf. C. Burt, 'Experimental Tests of General Intelligence', *Brit. J. Psychol.*, III, 1909, pp. 94–177: this paper was written while I was working as Sherrington's assistant at Liverpool, and it owed much to discussions of the problem with him. For a more detailed comparison of the results of factor analysis with theories about the organization of the brain, see Burt, 'The Structure of the Mind: A Review of the Results of Factor Analysis', *Brit. J. Educ. Psychol.*, XIX, pp. 100–11, 176–99, and refs.

The Reaction to Mental Tests. How then does the brain react to the various tests which the psychologist applies? In answering this question it will be sufficient to think of the human brain, like the rest of the body, as a machine operating according to the ordinary laws of physics and chemistry—but certainly, if it is a machine, it is by far the most complicated machine in existence, working on an exquisitely small scale. As with other machines we may regard its efficiency as depending upon (*a*) its structure and (*b*) the energy which runs it. Roughly, but not very precisely, these two aspects correspond to the 'cognitive' and the 'conative' aspects of human activity.

The cortex is a convoluted layer of tissue having an area of about 2,400 cm^2 (roughly 400 sq in.), but a thickness of only about 10·24 cm (1/10 in.). This implies a total volume of about 570 cm^3. But each square millimetre contains about 50,000 nerve-cells; and an average brain contains nearly 93,000,000,000 nerve cells.[2] Now this of itself makes Thomson's sampling theory extremely difficult to accept; for his theory requires that, when the correlation between two tests is high, each test must sample almost the whole of the nerve-cells in the brain (in Thomson's example between two-thirds and five-sixths); and it is impossible to believe that one test alone could sample as many as 70,000,000,000 cells.

So far as we know the nerve-cells in the brain that seem specially concerned with conscious processes do not differ fundamentally in their qualitative properties from nerve-cells elsewhere—those, for example, in the grey matter of the spinal cord which have no direct relation to consciousness. But the information about the outer world as it emerges in consciousness exhibits striking qualitative differences: sounds are different from smells or colours, and colours may be red, green, or blue, and so forth. But neurons going to, and through, and from the brain conduct only one kind of change—an electro-chemical wave travelling at about 300 feet per second. Thus information as received and transmitted by the brain displays no qualitative variations: it is, so to speak, coded in the guise of a set of similar signals differing solely in their number, their locality, and their mutual relations. We should therefore be able to handle it by purely quantitative methods. However, as is obvious from the vast numbers of items or events concerned, we are confronted with "a system comprising components whose interactions can only be described in terms of probabilities". Hence, in spite of the objections advanced by Dr. Taylor (this *Journal*, XI, pp. 133f.), a *statistical treatment* is both appropriate and essential (Sholl, *op. cit.*, pp. 62f.).

---

[1] This hypothesis was put forward fifty years ago in a paper describing correlations between tests for various intellectual processes. "If we ask how we are to picture the basis for a mental 'factor' or trait so general and yet biologically transmitted, I suggest we can identify it with the strength of a developmental tendency towards systematic structure in the central nervous system or brain. It is this that furnishes the possibility of an ever-changing adaptation to an ever-changing environment, and adaptation is the prime function of mind, that is, of attentive consciousness. A mind therefore will be most efficient where the inborn capacity for *complex organization* in the nerve-tissue . . . is richest" (Burt, 'The Experimental Study of General Intelligence', *Child Study*, IV, 1911, pp. 21-23).

[2] H. B. Thompson, 'The Total Number of Functional Nerve Cells in the Cerebral Cortex of Man', *J. Comp. Neurol.*, IX, 1899, pp. 113–140. This is the estimate still most frequently cited. Later investigators on the whole give rather lower figures (between 5×10^9 and 7×10^9 neurons): the chimpanzee has only about three-quarters that number, and *Cercopithecus* only about a quarter (G. A. Shariff, 'Cell Counts in the Primate Cerebral Cortex', *J. Comp. Neurol.*, XCVIII, 1953, pp. 381–400).

A single neuron operates on a power of about one or two thousand millionths of a watt; so that the whole brain operates on a power of between 10 and 20 watts.[1] A single nerve-impulse passing across a cell-junction produces a transient electrical charge equivalent to about 0·001 V lasting for about 0·01 sec or a little more. To cause a neutron to discharge, however, an impulse of nearly 10 times this voltage is required. By means of its branches a single neuron may impinge on a hundred or more neurons either in the neighbourhood or in more distant parts; but only about half-a-dozen will, as a rule, be in a state of critical excitability. In that case the cumulative effect of impulses arriving from a number of neurons at once will be to *increase the probability* that some or all these half-dozen will discharge in their usual all-or-none fashion. If then at each successive relay stage every neuron is in potential connection with a hundred more, it is clear that in a fraction of a second an incoming impulse may spread through the ramifying reticulations of the nerve-net to thousands of other neurons. The passage of a nerve-impulse in a 'neuronic arc' is therefore more like the passage of a wave front than of a simple current along a wire.

The Objections to a Theory of Random Sampling. No doubt in certain cases the course of the excitation may be to some extent a matter of chance. When you or I are meditating on a problem that calls for an imaginative solution, introspection will often suggest that 'chance' (or, as some psychologists prefer to say, 'the unconscious') is in the end responsible for a satisfactory answer; and work on trial-and-error processes fully confirms this inference. To some extent, therefore, the brain is undoubtedly a probability machine. I vividly recollect, while I watched Sherrington stimulating the cortex of some anaesthetized animal, how uncertain were the effects of stimulating precisely the same spot. As Sherrington puts it, "every increase in the number of links composing a nerve-cell chain seems greatly to increase the uncertainty of its reactions". Nevertheless, the uncertainty is never complete. Though the variations are at times bewildering, regularity seems on the whole the general rule.

It appears therefore impossible to accept the view that the *essential* nature of a test is to "activate a random sample of the nerve-cells of the brain". During states of sleep or inattention the diffuse and random activity of the cells, going round and round in reverberating circuits more or less in phase, would presumably set up the well-known ten-to-a-second alpha rhythm. An incoming sensory stimulus, however, would at once substitute a definite pattern of activity; and, according to the sense-organ from which they come, the impulses, still preserving their distinctive arrangement, will as a rule find their way to the appropriate regions of the cortex, and ultimately, if unimpeded, to what are the more or less appropriate motor centres. And when the outcome is successful, there seems comparatively little that could be put down to the sheer haphazard processes of random sampling.

Certainly, according to patients' reports, an electrical stimulus applied *direct* to the cortex does at times produce something rather like a random sampling of conscious effects—noises, coloured lights, odd skin sensations, differing unpredictably from time to time. Bang your ulnar nerve against the 'funny bone' in your elbow, and you get a random sample of all the possible sensations which the nerve could convey. But with the normal exploration of surfaces by the hand, the same nerve conveys a pattern of impulses, and you become aware of a significant tactile pattern corresponding closely to the pattern of the stimulating surface. And in much the same way, as I have argued elsewhere, an effective test of intelligence "operates by *selecting* neural bonds rather than by *sampling* them". Nor does it merely select; it organizes the bonds and the activities they mediate into systematic schemes of a complexity varying with the organizing skill of which the individua is capable. In short, to borrow Sherrington's phrase, the essential function of the mind or nervous system consists, not in its sampling action, but in its 'integrative action'.[1]

---

[1] 1 watt $=10^7$ ergs per sec or 33,000 ft. Lb. per min, i.e., 1/746 h.p. In terms of heat, 1 watt $=0{\cdot}057$ British Thermal Unit per min. If the power that runs the brain were used to boil water instead, it would need 3 hours to boil a pint; used for lighting it would operate only a very dim electric bulb to illuminate a passage.

Now for quantitative purposes we might as a first approximation treat complexity as a function of the number of possible connections. Accordingly, if there were no significant specialization in structure or function in the different regions of the brain, we might conceive the general factor of intelligence as measurable by the *average* number of possible connections per unit volume of the individual's brain. Since areas and layers do differ, I suggest it would be better to express it by the *minimal* number of possible connections per unit volume averaged over all the brain. The exceptional efficiency of particular processes might then be expressed by the *additional* number of possible connections in the different areas, i.e., the areas for vision, hearing, speech, etc.: these would appear as additive group factors.[2] Finally, the uncertainties involved in making the right connections could be most conveniently treated as forming an independent factor of 'unreliability', part of what the statistical psychologist calls the 'specific factor' (or 'uniqueness') of the test. 'Slips and accidents' of this type (often resulting from quite extraneous circumstances) are no more to be explained in terms of the 'sampling theory' or the 'principle of uncertainty' than the fact that Ben Walters' bullet "went wildly astray and hit the wrong chap". From the physicists' standpoint such 'slips and accidents' are quite as much the result of strict causal determination as the trajectory of the bullet. The uncertainty involved in Heisenberg's principle is of a very different type: it marks an issue which, so far as physical influences are concerned, is *entirely* undertermined.

The Uncertainty Principle. My allusions to the so-called uncertainty principle arose out of questions relating, not to 'intelligence', but to 'will'. Mr. Clarke, quoting Hebb and F. V. Smith, had argued at some length that "as a science, modern psychology assumes that all behaviour is strictly determined", and that "physical science, being based on strict causation, leaves no room for what is commonly termed 'free will'"; other writers, like Taylor and Hansel, have adopted much the same point of view. One or two, indeed, like Dr. Levy, do not wish to deny the uncertainty principle as a principle of micro-physics; but they maintain, as he does, that the neural processes of the brain are on too large a scale to be affected by the operation of that principle, even if we give it our assent.

The problem of size can be settled quite simply. What are popularly called 'nerve-cells' (the cell bodies of neurons) vary greatly in their bulk. The giant pyramidal cells measure as much as 60 to 80 microns in their diameter; but some of the smaller types measure only 4 to 6 microns. But here we are concerned, not with nerve-*cells*, but with nerve-*junctions*. The fibre-like branches of the nerve-cells subdivide into twig-like filaments; and in the spinal cord these filaments are seen to end in little buttons or knobs (*boutons terminaux*) which are applied to the cell bodies of the adjacent neurons or to their branches. There is no structural continuity between one neuron and another: they are apparently separated by a film that may perhaps consist of just one molecular layer. On the surface-phenomena which thus arise, all the properties of the cell-junction or 'synapse' essentially depend. In particular this tiny gap and the chemical substances that occupy it, interpose a varying resistance to the passage of nerve impulses; and, like a valve, permit it to travel in one direction only.

---

[1] *Brit. J. Psychol.*, XXX, pp. 87f.

[2] This highly simplified scheme will fit a simple verbal test of intelligence, such as Terman's vocabulary test, quite neatly. But such tests ignore the process of efficiently *combining* connections into an organized whole that is so distinctive a feature of the higher types of intelligent process. Nevertheless, that is not so serious as it might seem. In general the brain can only work with an increased number of connections when it is able to organize them into a system: try tapping with the hand and feet while nodding the head; it is almost impossible to effect the connections required in a random order; inevitably they fall into a rhythm. However, to make sure of testing the capacity for organization I used relational tests in which multiple connections were involved, e.g., tests of implicit or explicit reasoning, and then counted the number of correct combinations made in carrying it out. Such a test of course departs still further from the process of 'taking random samples of nerve-cells': (cf. *J. Exp. Ped.*, I, pp. 93f., *Brit. J. Educ. Psychol.*, XIX, pp. 176f.)

In the cortex the 'synaptic knobs' either fail to develop or fail to stain. Nevertheless, as measured in the spinal cord they give some indication of the scale of size with which we are here concerned. The diameter of a knob averages about 1 micron; and its mass must therefore be of the order of about 5×10^{-13} gm.[1] Let us now apply Heisenberg's equation.[2]

$$\Delta x \,.\, \Delta p \geqslant \tfrac{1}{2}(h/2\pi),$$

where x denotes position, p momentum, and h is Planck's quantum of action ($6{\cdot}55 \times 10^{-27}$ erg/sec). A little calculation shows that, with a small body having the weight of a synaptic knob, there would be an uncertainty in the actual position of about 20 Ångström units in 1 sec, or about 2 Ångström units in 0·01 sec. Now the gap between the terminals of one neuron and the surface of the next must be only 200 Ångström units across. Evidently an uncertainty of the order indicated might easily introduce an uncertainty into the strength needed by the incoming nerve impulse in order to leap across the gap.[3]

The phrase 'neuron-arc', used by Thomson and other writers, conveys the notion that the cortical associations or 'bonds' are effected by end-to-end chains of *single* nerve-cells, each forming a separate track or pathway. This notion perhaps springs partly from the simplified accounts of the 'sensori-motor arc' that we find in such books as McDougall's *Physiological Psychology*, or Watson's *Psychology from the Standpoint of a Behaviourist* (cf. Figures 25 and 46), and partly from the ever-popular comparison of the nerves to telegraph wires and the brain to a central exchange. But if that were how the brain was constructed, even 90 thousand million cells would not suffice to mediate the highly intricate and highly variable patterns of human behaviour: a simple calculation will show that the mere act of reading this paper would call for more cells than that.

However, what we now know of the organization and functioning of the cortex points to a very different mode of working. As we have just seen, the structure of the cortex reveals, not a system of distinct and specific channels, running from the input-areas to the output-areas, but an intricate three-dimensional, or rather multi-dimensional network; and the most significant feature of every nerve-cell is the way in which its axon (the outgoing fibre conducting impulses away from the cell-body) sooner or later subdivides into numerous branches, and thus makes contact not with one neuron, but with a number. To illustrate the possibilities which all this ramification introduces consider the following alternative arrangements. Let A1, A2, . . . , B1, B2, . . . , C1, C2, represent cell-bodies, and suppose for simplicity that the axon from each subdivides into four branches only, and that impulses from at least two branches are needed to make the next neuron discharge. Impulses arrive at A3, A4, and A5, which thus are ready to fire. Now let us suppose that, owing either to past activity or to impulses arriving from some other cells, the synapses at B2, B3, and B4 have their thresholds temporarily lowered: then the incoming stream of impulses will be passed on in the direction of C1, C2, and C3 (Figure 1). If, however, the lowest thresholds are at B4, B5, and B6, then the stream will flow in the direction of C5, C6, and C7 (Figure 2).

---

[1] R. A. Haggar and M. L. Barr, 'Quantitative Data on the Size of the Synaptic End-Bulbs in the Cat's Spinal Cord', *J. Comp. Neurol.*, XCIII, pp. 17–35.

[2] Cf. 'Quantum Theory and the Principle of Indeterminacy', this *Journal* XI, 1958, pp. 78f.

[3] A micron is a thousandth of a millimetre. An Ångström unit is one ten millionth of a millimetre. The electrical effect produced by a nerve impulse arriving across a synapse is extremely small and transitory—equivalent to about 0·001 V and lasting between 0·01 and 0·02 sec. An excitation having at least ten times this voltage is needed to trigger off a discharge from the post-synaptic neuron: after firing, several milliseconds are required before it is in a position to fire again.

I should perhaps add that it is now generally believed that the transmission of a nerve-impulse across the synapse is effected by the rapid liberation of a chemical substance as a result of the arrival of the pre-synaptic impulse. The substance (possibly acetylcholine) is presumably stored in the nerve-terminals; its diffusion and momentary combination with the substances of the intervening film break down the resistance, generate a potential, and thus set up a wave of depolarization in the post-synaptic neuron: while it diffuses, it is rapidly destroyed by an enzyme (e.g., cholinesterase): (cf. W. Feldberg, Gegenwärtige Probleme auf dem Gebiet der chemischen Übertragung von Nervenwirkungen, *Arch. Exp. Path. u. Pharm.*, CCXII, 1950, pp. 64–68).

FIGURE 1

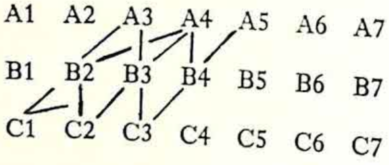
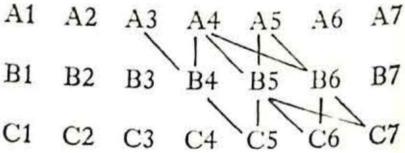


FIGURE 2



This yields a picture of the brain and its working very different from the behaviouristic theory that human conduct is just a matter of stimulus and response, and the brain itself just an organ for conditioned reflexes, "the whole acting like a set of telephone wires with automatic switches". As we now see, there is, or there need be, no specificity about the neural channels. Streams going in very different directions at slightly different times may flow like waves across the same groups of neurons. Further two or more streams may converge and coalesce; or a single stream may conceivably divide into several. Some streams or their offshoots may fade out; others may provide a feedback; and others again may go round and round in circles passing and repassing over the same set of nerve-cells.

It was largely on these grounds that I suggested the use of a 'hypothetical model derived from hydrodynamics', and of concepts, terms, and analogies borrowed from the study of moving streams and currents. Applied to states of nervous instability, words like 'turbulence' were used, not as vague metaphors, but as technical terms having much the same meaning as they are given in the theory of fluids generally.[1] Here are a series of problems and a practicable mathematical technique to which both the quantitative psychologist and the clinical psychologist can usefully turn their attention.

A. THE BRAIN AND PERCEPTION

Perceptual Space, Physical Space, and Factorial Hyperspace. The physiologist studying the nervous system assumes that the things that he examines—the brain, its constituent cells and fibres, the electro-chemical impulses propagated along the nerves—as well as the apparatus he uses to examine them—are all material objects existing independently of his own perceptions and thoughts, in a three-dimensional physical space. Within the brain the nerve-impulses themselves, as we have seen, form a spatio-temporal pattern. The sensory percepts of the owner of that brain—the coloured patches that form his visual field, the tangible shapes he feels with his hands—also form spatio-temporal patterns. And these two patterned schemes—the neural and the sensory—are related in various ways to the pattern of events in the physical space surrounding the owner's brain and body. They

[1] When a fluid, whether liquid or gas, flows with a moderate and even velocity through a tube or other channel having a smooth, uniform and relatively narrow bore, the flow is said to be 'streamline', i.e. the velocity contains no components transverse to the main direction of the stream. If the bore is irregular, then, as the pressure at the inflow is increased, a critical point is eventually reached at which the type of motion changes; eddies and minor currents are formed, and as a result the flow becomes erratic both in rate and in direction: such a flow is described as 'turbulent'. This conception, as it seemed to me, served very well to describe the condition underlying certain states of intellectual confusion and emotional disturbance. Dr. Gregory has adopted a similar model to account for what, to the clinical observer, seems like a qualitative change in the behaviour of neurotic and psychotic individuals under certain conditions of pressure and stress (cf. this *Journal*, XI, pp. 211f. and refs.).

appear in fact to be initiated by stimulating sources located within that space; and they form his chief means of discovering what is happening there.

Dr. Levy asks whether it is possible to regard these three spatial schemes as parts of one all-embracing space or rather space-time; and he suggests that this all-embracing space-time may be regarded as a kind of " factorial space ", like that displayed in my diagram in an earlier issue of this *Journal* (VIII, p. 110). Of the three dimensions there depicted one explicitly represents ' time ': the second—a non-personal dimension allocated to ' tests '—really represents, he says, ' a projection (on a single line) of physical space and the events within it '; and finally the third, labelled ' persons ', represents (in linear projection) the various ' mental spaces ' of individuals concerned. Now physical space is three-dimensional; and each ' mental space ' is, in his view, also three-dimensional. He thus arrives at a ' psychophysical factor-space '—a hyperspace which, like that described by Dr. Smythies,[1] has $3(m+1)$ dimensions, m being the number of individuals. The study of this psychophysical space, he contends, is essentially a task for the statistical psychologist, and, if successful, might lay the foundations for a " psychophysical cosmology that would take the place of traditional metaphysics ".

Incidentally, he tells us, this mode of interpreting factorial results comes closer to that adopted by Godfrey Thomson than to that favoured by Thurstone, Spearman, and myself; and once again illustrates the fruitfulness of the sampling theory.

The Ambiguity of Space. Dr. Levy does not himself develop any detailed arguments in support of these views, but refers the reader to Dr. Smythies. His extension of Dr. Smythies' conclusions, however, seems to bring out very clearly what I hold to be a manifest fallacy common to both. In my view both of them seem to have taken a highly ambiguous concept ' space ' and have applied it to very disparate sets of phenomena; and by this means they claim to have shown that all these phenomena are part of a single all-inclusive whole, which is also spatial.

In factor analysis the use of geometrical terms and spatial concepts is no more than an expository device. It furnishes a convenient language. Two or more variables that vary together can be represented either by the algebraic symbolism of numbers and letters or by the geometric symbolism of points, lines, planes, and angles. The psychological factorist freely exploits both methods: but, when he makes use of the geometric symbolism and talks of a factor ' space ', he does not for a moment intend the ' space ' to be regarded as a concrete entity.[2] In any case the diagram mentioned, like Thomson's spatial charts, would need a new interpretation before it could be identified with a psychophysical hyperspace: for instance, what Dr. Levy calls the ' non-personal dimension ' describes the common characteristics, not of the physical universe, but of the persons tested (or, in Thomson's illustrations, their abilities).

---

[1] J. R. Smythies, ' Three Classical Theories of Mind ', *J. Soc. Psychic. Res.*, XL, 1960, pp. 385–96, and *id.*, *Analysis of Perception*, 1956, pp. 29, 129f.

[2] From what he has said elsewhere it is clear that Dr. Levy is well aware of the symbolic nature of the terms used by the factorist; and possibly he would reply that the terminology used by physicists to describe physical space, and the terminology used by philosophers to describe perceptual space are at bottom themselves only symbolic. But in that case the whole argument proves nothing. I might just as well argue that because the symbol p is used by a geographer to denote a point, by a statistician to denote a probability, and by an economist to denote a profit, therefore points, probabilities, and profits are really the same.

The preference for geometrical procedures is well marked with those factorists who, like Thomson and Thurstone, are avowed visualizers; a preference for algebraic procedures is more usual with those whose visual images are, like my own, evanescent and dim. Like Lord Kelvin, the visualizer commonly feels he has mastered a problem when he has hit on a way to picture all the details in a visible model. I cannot help suspecting that Dr. Smythies' descriptions of his own experiences suggest that he at any rate possesses vivid visual imagery; and I wonder whether Dr. Levy may not enjoy the same advantage.

The Representative Theory of Perception. Dr. Smythies' views may best be summarized in his own words. The hypothesis he wishes to support is as follows: " Sense-data and images are spatial entities distinct from physical objects, and bear both temporal and causal relations and higher dimensional spatial relations to physical objects ".[1] He claims two special advantages for his theory over " the traditional dualism ". In opposition to the Cartesian doctrine, which makes mind an unextended and non-spatial substance and so lands us with the anomaly of psychophysical interaction, the new theory holds (i) that " mind and body form a single organism of which they are different *spatial* parts " and (ii) that we are thus enabled to give " an intelligible account of sense-data and the rôle they play in perception ". For if mental events are themselves ' spatial entities ' in the same hyperspace as the external stimuli or the nerve processes in the brain, then, it is maintained, we can validly account for those entities in terms of cause and effect. This in turn will justify us in treating our sense-data as supplying a representation or photographic copy of the events in the physical universe. And on the basis of this interpretation Dr. Smythies concludes that " the mind-brain problem can now be solved by advances in neurophysiology ".

Dr. Smythies works out his " modified version of the ' representative theory ' of perception " chiefly in terms of visual sense-data. But there are other kinds of sense-data—sounds and smells, for example—which possess few, if any, of the spatial characteristics he specifies, and ' mind ' includes many types of ' conscious content ' besides sense-data—pleasure, happiness, anger, imageless thoughts—which have no spatial characteristics whatever. Hence the *whole* mind–brain problem cannot be solved in this way.

Further, the physicists who have accepted the traditional notion of physical causation have almost invariably assumed that the events which act as causes and the events which ensue as their effects are and must be events in the *same* space (or rather the same space-time). Dr. Smythies adduces no evidence whatever for supposing that physical events occurring in ' public ', ' objective ', or ' physical ' space could cause mental events in a ' private ', ' subjective ', and ' non-physical ' space. The mystery of psychophysical interaction therefore seems no nearer solution. And of course the whole notion of causation has of late been vigorously challenged, particularly as applied to the transition from processes in the brain to changes in consciousness.[2]

Dr. Smythies seems to suppose that he is explaining how an observer's private perceptual space could be described in terms of the spatial relations recognized by scientists in their geometrical accounts of physical space; but what in fact he is really doing is to interpret, or rather to picture, the scientist's conception of physical space by analogy with the perceptual space of binocular vision. His aim is to show that the visual field is really a copy or representation of the physical field; but he does this by assuming the physical field is just a replica of the visual. Thus, he repeatedly compares the resemblance between "the perceptions in our minds" (as people say) and "the events outside our minds" with the resemblance between a moving picture as seen on a television screen and the actual events as scanned by the television camera on the spot. He even claims that "the mechanisms of visual perception *are* television mechanisms" (*loc. cit.*, p. 42, his italics); and he warns us against supposing that the comparison is "*only* an analogy".

---

[1] This he calls Theory II. Theory I is expressed in similar terms, except that it denies that *spatial* relations connect sense-data and objects. The latter is the theory advanced by H. H. Price (*Perception*, 1950) and more recently by H. Kuhlenbeck (*Brain and Consciousness*, 1957). A final decision between the two hypotheses, says Dr. Smythies, must await ' experimental demonstration '; and, since the entities concerned (including the sense-data) are assumed to be spatial, the demonstration will require the techniques of ' mathematical *physics* ', not, as Fechner and his followers assumed, of mathematical *psychology* or psychophysics. Meanwhile, Dr. Smythies' criticisms of Price's views plainly imply that he himself favours ' Theory II '. It may be added that the idea of a " *space-like* whole of more than three dimensions in which both sensa and scientific objects literally have places " had already been put forward as one possible hypothesis by Broad in his *Scientific Thought* (1927, pp. 392f., 543f.).

[2] E.g., by A. J. Ayer among contemporary philosophers (*The Foundations of Empirical Knowledge*, 1940, pp. 220f.). Dr. Smythies suggests that, with his hypothesis, those who still see difficulties in his theory of psycho-neural interaction could substitute a theory of a psycho-neural parallelism. But here, I suspect, he would find most modern neurologists against him: Sherrington, for instance, declares that " although the physical basis encroaches more and more on the study of the mind, there are mental events which seem to lie beyond any physiology of the brain "; and Adrian observes that there seems to be " an important part of the picture—the part that deals with the mind—which can never be fitted into our picture of the brain " (ap. P. Laslett, *The Physical Basis of Mind*, 1950, pp. 3, 5). The question is largely verbal; and many of the puzzles may be evaded by substituting the language of information theory for that of causal theories.

There are several difficulties here. On the physiological side, the stream of electro-chemical nerve-impulses, as is evident from the description just given, bear no resemblance to a modulated radio-wave; nor is there any mechanism in the eye or brain to perform the scanning that is done by the television camera. On the psychological side, our reflective reconstruction of the physical world is based on a quasi-inferential procedure—namely, on what Spearman would call an 'eduction of correlates'. We have no means whatever of directly comparing the actual physical events with the alleged mental copy in the way we can compare them with the copy on the television screen. And in any case I should have thought the differences were far more startling than the similarities. The physical world is devoid of all secondary qualities (unlike the mind, the ordinary television camera leaves out the colours instead of putting them in). And if, as the modern physicist maintains, the events in the physical world consist of particles gyrating round each other and emitting waves of probability that traverse space-time in geodesics, then none of this finds any representation in our visual field.

However, if we waive these rather dubious analogies, what Dr. Smythies' theory really comes to is this: the individual's 'personal space' as studied by the psychologist and the 'scientific space' postulated by the physicist exhibit no differences in kind: they are just "two different sections of an n-dimensional space, not two *kinds* of section". But I doubt if either the psychologist or the physicist would agree with his versions of space as each of them understands it.

(i) To begin with, every individual experiences not one sensible space, but several. Tactile space is different from visual; auditory space (as the introspections of those born blind clearly show) is different from either;[1] when their eyes are open, many people (I for one) locate their visual images in a space which is quite unrelated to any space that is actually perceived. The part played by the proprioceptors is more obscure. Without these, and particularly without the contributions from the inner ear, we should have a very poor notion of direction and orientation: and what meaning can 'up' and 'down' have for an astronaut in a windowless capsule when deprived of the sensation of weight? Finally, the temporal integration of kinaesthetic sensation seems essential for the recognition of distances traversed.

Certainly it is quite unsafe to identify the individual's 'private' or 'subjective' space with his 'visual space'. This in any case is not a 'space' (which is an infinite continuum) but a 'field' (a very restricted type of continuum). In a factorial study carried out by Miss E. A. Hudson the largest contribution to the general factor of subjective space came, not from visual perception, but from the perception of movement; verbal descriptions came second; visual imagery third; visual perception only fourth. Moreover there were wide individual differences. A man's private or 'sensible' space is therefore, not a simple three-dimensional percept but a highly artificial construction, quite as much conceptual as perceptual. If the man were a physicist, it would not be a *section* of the same whole as physical space; it would be identical with physical space as he conceives it, for that is also a highly artificial concept.

---

[1] It is a little surprising that, since our first ancestors emerged from the sea and became land-animals, we have evolved no new sense-organ for the perception of space or of the energies which operate in space.

(ii) As we have seen, Dr. Smythies' own conception of 'scientific space' seems to resemble a psychological construction based on the space of binocular vision rather than the sophisticated concept of the modern physicist. A questionnaire addressed to students (including students of physics as well as of psychology) indicates that the majority of them picture it, much as Plato did,[1] as a kind of huge box with no sides or bottom, or as a vast room whose ceiling, walls, and floor are infinitely remote, a receptacle in which all physical objects can be contained, but possessing no physical properties or concrete structure of its own. This in fact was the notion of Absolute Space adopted by Newton and most of his contemporaries. But it is not the conception of the modern physicist. For him the word space no longer suggests a great container—an actual three-dimensional framework in which physical substances are located. The term has become simply a convenient name for certain highly abstract relational properties—notably those of order—which characterize the spatio-temporal whole we call the physical universe.[2]

Since subjective or sensible space and objective or physical space are both parts of a single spatial whole, Dr. Smythies argues that "sensa and physical objects must bear spatial relations to each other". Nevertheless it is hard to see what advantages are gained either by this hypothetical union of the sorts of space or by the somewhat unexpected inference that he draws. I can see a certain plausibility in unifying subjective time with objective time: e.g., in assuming that events in my own conscious life-history form part of the wider series of events that make up the life-history of the universe; something of this sort seems implied when I talk of waking at the proper time to watch the sun rise at the solstice. Yet even here I doubt whether this is the right assumption. However, Dr. Smythies tells us that with his spatial model "the visual field is able to represent a portion of the physical world", much as a mirror will re-present an object when the two are (so to speak) face to face in the same room. Apparently he supposes that sense-data can only yield knowledge of physical objects, if they resemble them.[1] But a mirror-like resemblance is an unnecessary and indeed an improbable requirement. What is needed is merely that the two shall approximately correspond.[2] Without some such correspondence our behaviour would fail to effect its aims; and neither we nor our species could expect to survive. But the correspondence need only be a correspondence of order. The philosopher, and still more often the philosophizing physiologist seem prone to forget that natural selection does not cater for scientific curiosity about the universe: its concern is merely that every life should gain a living.

---

[1] πάσης γενέσεως ὑποδοχή ('a receptacle for all created things') Plato, *Timaeus*, 49 A cf 51 A.

[2] This view is as old as Newton's great contemporary Leibniz: "There is no actual empty space or actual empty time. Space is merely relational as time is. I hold it to be an order of possible coexistences" (*Troisième Lettre à Clarke*; cf. *Réplique aux Réflexions de Boyle*, 1702, Erdmann's edition 189b). Leibniz's doctrine, however, was emphatically rejected by Kant: "Der Raum ist kein diskursiver, oder, wie man sagt, allgemeiner Begriff von Verhältnissen der Dinge" (*Kritik der reinen Vernunft*, 1, 1ster Abschn., 2(4)). In Britain, even before the advent of Einstein's relativity, the relational view was supported by Ward among psychologists and by Russell among philosophers (cf. B. Russell, *Critical Exposition of the Philosophy of Leibniz*, 1900, and earlier papers). Today, of course, the relations postulated would be relations, not between *points* in space *or* time, but between *events* in *space-time*. However, until quite recently, most British psychologists, physiologists and philosophers seem to have preferred the Newtonian view. Continental writers have been greatly influenced by the views of Hilbert. For him points may be denoted by *any* undefined elements that obey certain axioms; lines and planes are undefined *classes* of points; dimensionality is not the property of class as such, but of *order* in a class. Moreover, most contemporary physicists, I imagine, would hold that the topological structure of the physical world (i.e. of what Dr. Smythies calls 'scientific space') is independent of measurable magnitudes. Magnitudes appear with the use of coordinates and the choice of a metric, but they are eventually eliminated as a result of considering only those properties which are invariant when the coordinates and the metric are changed. Thus what the physicist is really concerned with is primarily the *logical* structure of the physical world; and what the layman thinks of as its 'mathematical' features—the determination of angles, distances, and various other quantities—are introduced merely for the convenience of proof.

Professor Broad brings all this out very neatly by suggesting that, instead of picturing spatial structures in visual terms, we should think of the relations involved as relations between members of a huge family: the various spaces of different persons would then be expressed by the relations subsisting between members of certain sub-families, i.e., of branches of the wider family (cf. *Scientific Thought*, pp. 544f.).

Correspondence can exist without both corresponding structures being spatial. Even if they were, spatial relations alone could not guarantee what is required. It is the correspondences we must investigate. Since they are never exact enough to amount to strict isomorphism, a correlational technique would seem to furnish the simplest means for most investigations. Here again there are obvious and urgent problems for the statistical psychologist to attack.

B. The Brain and Learning

Mathematical Models. There have been several attempts to construct mathematical models simulating a nervous network with completely random interconnections, supplemented as a rule by devices for reinforcing appropriate connections and eliminating inappropriate connections. Some of the simplest have been embodied in actual working models. The most ambitious perhaps is the ' Perceptron '.[3] From all of them the conclusion to be drawn is much the same: a mechanism based on a completely random network learns extremely little extremely slowly, and needs an extremely large number of elements to accomplish even the simplest type of discriminative reaction. Whenever the task is at all complex, some degree of non-randomness seems essential, though a certain degree of randomness renders behaviour more elastic and so more adaptable; and where learning entails in addition some type of problem-solving, then, it would seem, no computing machine working on an entirely random network can succeed[4].

The Neurological Processes. Far less is known about the neurological basis of learning than Dr. Levy appears to assume. Both Eccles and Kinorski agree that the hypothetical physiological concepts postulated by Pavlov—irradiation, induction, internal and external inhibition, and the like—"have no neurophysiological or neuroanatomical meaning".[1] As we have seen, the anatomical arrangement of neurons in the cortex would undoubtedly allow impulses to keep circling around, by a kind of reverberative activity, in closed self-re-exciting chains;[2] and Rashevsky, Young, Householder and Landahl, and others have attributed both learning and retention to these continuously circulating patterns.[3] Such a hypothesis might account for short-distance memory; but it can hardly account for more permanent storage, for it is difficult to suppose that all these circulating patterns persist even during deep or complete anaesthesia.

---

[1] The half-unconscious aim seems to be to account for cognition without invoking any sort of specifically cognitive or mental relations. The more familiar causal relations, spatial relations, and relations of similarity are each in their turn treated as capable of doing the job. I hold that it is impossible to dispense with a cognitive relation *sui generis*—a relation of direct awareness, peculiar to those organisms which possess consciousness. But I have set out my arguments more fully elsewhere and need not repeat them here.

[2] I shall not attempt a formal definition of ' correspondence ' here. I regard it as a generalization of what in matrix algebra is called ' equivalence '. An illustration may help. The notes we see on the staves of a printed orchestral score form a spatial arrangement of spatial shapes. They *correspond* with the musical sounds as heard, but in no way *resemble* them. The sounds are not spatial; still less are their relations to the corresponding notes spatial relations. In point of fact, if the contents of our ' private sensory space ' formed a complete facsimile of the physical world it would be most confusing: fortunately our sense-organs select and simplify those crucial features to which it is imperative we should respond if we are to survive.

[3] Cf. N. Rashevsky, *Mathematical Biophysics*, 1948. W. S. McCulloch and W. Pitts, 'A Logical Calculus of the Ideas Immanent in Nervous Activity ', *Bull. Math. Biophys.*, V, 1943, pp. 115f. W. R. Ashby, *Design for a Brain*, 1952. W. Grey Walter, *The Living Brain*, 1953. F. Rosenblatt, ' The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain ', *Psychol. Rev.*, 1958, pp. 386f.

[4] Cf. R. M. Friedberg, 'A Learning Machine ', *IBM J. Res. and Devel.*, II, 1958, pp. 2f.

On the whole the hypothetical model which seems most convincing is that schematically outlined long ago by McDougall: it assumes overlapping of impulses, synaptic facilitation, and a permanent lowering of synaptic resistance.[4] There is now a good deal of experimental evidence to support these several assumptions;[5] and if this hypothesis be accepted, then it means that there must be *some* specificity in the neural pathways. Dr. Levy contends that the statistical studies of learning which Foley and I carried out themselves really imply a hypothesis of random sampling. I would prefer to describe our basic assumption as depending not on a sampling theory, but simply on a stochastic theory; it postulates, not complete uncertainty at the initial stage, but only a partial uncertainty. What, as it seems, adaptation requires is not that the neural responses should be entirely random, but that they should include some measure of freedom. For the rest the process is still one of organization rather than of sampling.

C. The Brain and the Will

The Nature of Will. Dr. Levy complains that, even in the discussions with Mr. Clarke and his deterministic supporters, I have "rather shirked the problem of the will as such". This was partly because 'will' is such an elusive concept, and partly because in considering the relations between brain and mind, the study of the cognitive problem is logically prior to the study of the conative and likely to prove far more informative. Many psychologists have denied that anything like 'will' could affect the brain; but no one can deny that the brain affects cognition.

If we are to discuss the problem of 'will', we must first agree how the term is to be understood. For the purposes of the present argument I shall take the word to mean a conscious, non-physical agency whose existence and activities escape detection by the most delicate instruments that the physicist has been able to devise.[1] With that definition there would seem to be but two ways in which we could become aware of its working. (i) We may have a direct intuitive apprehension of its action. I myself am inclined to believe that this is true. No doubt the alleged experiences of a 'conative effort', as felt, for instance, when we endeavour to lift a heavy weight, consist in the main, of kinaesthetic sensations. But in wishing, yearning, wanting, and striving to put my wants into effect, I seem to have a direct awareness—even when no overt movement is made—of a genuine conative tendency; and this experience appears to supply the chief source for the anthropomorphic conception of force. However, these rather questionable introspective pronouncements can at most only be cited as confirmatory evidence, not as conclusive. (ii) But secondly, the brain itself may act as a detector. In that case the evidence for our belief in the efficacy of will would be indirect—an inference from the way in which voluntary effort modifies, or seems to modify, our bodily reactions. How far this conception of the brain is permissible or probable is a question to be answered in the first place by the physicist and physiologist; and it is to them the philosopher or psychologist must turn before offering views of his own.

---

[1] J. Konorski, *Conditioned Reflexes and Neuron Organization*, 1948; cf. *id.*, 'The Mechanisms of Learning', *Symp. Soc. Exp. Biol.*, IV, 1950, pp. 409f.

[2] Cf. R. Lorente de Nó, 'Studies on the Structure of the Cerebral Cortex', *J. Psych. Neurol.*, XLV, 1933, pp. 381f. and XLVI, 1934, pp. 113f. B. Delisle Burns, 'Some Properties of the Cat's Isolated Cerebral Cortex', *J. Physiol.*, CXI, 1950, pp. 50f.

[3] J. Z. Young, *ap.* G. R. de Beer, *Essays Presented to E. S. Goodrich*, 1938, pp. 179f.: A. S. Householder and D. H. Landahl, *Mathematical Biophysics of the Central Nervous System*, 1945. Everyone is aware how a tune or a phrase will "keep going round and round in one's head"; and no doubt iterative processes of this sort, both conscious and unconscious, help to keep alive a 'short distance recollection', while the 'long distance memory' is being driven home.

[4] W. McDougall, *Physiological Psychology*, 1905, chap. VIII. Compare his diagram on p. 143 with Eccles' diagrams, *op. cit.*, pp. 220 and 222. The chief difference is that Eccles is thinking in terms of streams or patterned 'columns' of neuronal activity rather than in terms of single neurons. But that is a modification McDougall would willingly have accepted. It was partly for this reason that I formulated my own arguments in terms of matrix notation and factor analysis: see also the recent work of L. R. Tucker, 'Determination of Generalized Learning Curves by Factor Analysis', *Psychol. Dept. Univ. Princeton*, 1960; and R. A. Weitzman, 'The Performance of Rats and Fish on a Probabilistic Learning Problem', *Educ. Testing Serv. Rep., Princeton*, 1959.

[5] There is, for example, direct evidence to show that 'use' leads to increased functional efficiency in the synaptic junctions and that 'disuse' leads to diminished function: cf. J. C. Eccles and A. K. McIntyre, 'Plasticity of Mammalian Monosynaptic Reflexes', *Nature*, CLXVII, pp. 466f., and *op cit. sup.*, pp. 203f.

The Physical Possibilities. Among physicists Eddington is the only writer who has examined the issues in some detail. He has at different times advanced three possible hypotheses.

(i) In what I will call his pre-Heisenberg writings he put forward his own version of the idealistic type of monism that had become increasingly common in this country from the days of Mill and Mach. "To state it crudely" (he says) "the stuff of the world is mind-stuff, . . . and within the world the main cleavage is, not between the material and the mental, but between the metrical and the non-metrical". Physical science deals only with the metrical aspect; mental attributes may be purely qualitative and are therefore non-metrical. Volition he held to be "genuine"; but he considered it to be "one of the most dubious points in the theory of the interaction of mind and matter".[2]

(ii) A few years later, however, he tells us that "the recent revolution which has expelled determinism from present-day physics . . . is sufficient to justify a reorientation of our attitude towards the problem [of free will]". And in this he has been followed by a number of younger quantum theorists. Yet "though the door of freedom is opened, it is not flung open wide". Overt bodily behaviour depends "on a kind of trigger action; a minute physical change in a nerve may release a vast amount of chemical energy pent up in the muscles". If therefore we accept the uncertainty principle (as most physicists do) it becomes conceivable that, within the very narrow margin which that principle allows, mind might behave as a kind of psychokinetic mechanism able to produce these "minute physical changes within a nerve".[3]

(iii) Nevertheless, after some hesitation, Eddington himself rejected this hypothesis for reasons to be discussed in a moment; and his last hypothesis turned on a distinction between the "uncorrelated or random behaviour" of particles in general as postulated in contemporary physics and the orderly and correlated behaviour of particles in those organized portions of matter which he considered to be in immediate relation with mind.[1] I should fully agree that orderly behaviour is a distinctive characteristic of efficient mental activity and that this is in marked contrast with the physical tendency to ultimate disorder, as expressed in Boltzmann's formulation of the second law of thermodynamics ("the total entropy of the universe is increasing and tends towards a maximum which will correspond with complete disorder among the particles it contains"). But this introduction of 'order' (or, as I would rather call it, organization) appears to be a description of the *nature* of mental activity rather than of the manner in which mental activity operates on the physical processes within the brain. Hence it need not be an alternative to the second hypothesis but rather a consequence of it.

---

[1] I should not agree that the brain is in any sense an 'organ of will'. In origin, I admit, it arises as a rudimentary outgrowth concerned essentially with the task of improved motor adjustment: i.e., it is primarily an organ of practical behaviour rather than of cognitive understanding. As an integrative mechanism its chief function is to undertake the complex organization needed for *doing one thing at a time*. To this the mental counterpart is in my view 'attention' rather than 'will'. As Sherrington puts it, "a 'willed' act is but the culmination of attention" (*Man in His Nature*, p. 270). However, Dr. Levy might retort that attention at its highest is willed attention, and involves a non-mechanical choice. So here I retain his key-term.

[2] A. S. Eddington, *The Nature of the Physical World*, 1928, pp. 275f. As he explains, however, his chapter on causation and volition had to be largely re-written owing to the fact that in 1927, less than twelve months before publication, "quantum theory definitely entered on a new phase" with the announcement of Heisenberg's principle of uncertainty.

[3] A. S. Eddington, *New Pathways in Science*, 1935, pp. 88f.

The chief argument which led Eddington to reject the second hypothesis (which so many other quantum theorists have preferred) may be summed up as follows. He "conjectures that the smallest [neurological] unit in which the physical effects of volition could have their origin must contain billions of atoms; and if such a unit behaved like an inorganic unit of the same mass, the indeterminacy would be insufficient to allow appreciable freedom". Now the lightest known atom (an atom of hydrogen) has a mass of $1{\cdot}6 \times 10^{-24}$ gm. And if Eddington is using the word billion in its English sense (a million million or 10^{12}), then a billion atoms of hydrogen would have a mass of $1{\cdot}6 \times 10^{-12}$. Now this is very much larger than the mass of a synaptic knob (5×10^{-13} gm). Hence it seems clear that Eddington was assuming that the smallest functional unit would be the nerve-cell. But if, as was argued above, we must regard the synapse or synaptic knob as the smallest functional unit, then his difficulty disappears.

The Physiological Problem. What then does modern neurology make of this suggestion? At this point we must plainly turn from the physicist to the physiologist. Professor Eccles, himself one of Sherrington's ablest collaborators, has discussed the issue at some length. As we have seen, the passage of a nervous impulse involves, not one, but a number of synaptic junctions. And in such a case, as Professor Eccles points out, if these minute influences were exerted at a large number of points consistently and simultaneously, the effects would be "rapidly integrated in the form of a changed spatio-temporal pattern of activity within the neuronal net". Thus, in the cortex as pictured by contemporary neurologists there unquestionably exists "a mechanism that could effectively amplify these minute effects, provided they form a 'meaningful' pattern" (i.e., presumably, a self-consistent and convergent pattern), "and are not random".[2]

We may therefore imagine that in certain circumstances the brain could act as a detector and amplifier of the minute effects exerted by the 'will', even when no known physical instrument is sufficiently delicate to detect them. But for this purpose certain conditions must be fulfilled.

(i) First of all, a certain number of neurons must already be at or near the critical level of instability. The influence of the 'will' would then be to increase the probability that some of these neurons will discharge. When the activity of *all* the cortical neurons sinks to a fairly low level, as in states of deep sleep, anaesthesia, and the like, the brain fails to respond to minute influences of the 'will', just as it fails to transmit impulses coming from the sense-organs unless they are exceptionally intense; and in these exceptional mental states the low level of activity throughout the brain can be readily demonstrated by the electro-encephalogram. However, for the manifestation of consciousness or will only a part of the brain need be active at a critical level: the rest will supply merely background activity. And apparently almost any part—almost any set of interrelated cells—will serve.[1]

---

[1] A. S. Eddington, *The Philosophy of Physical Science*, 1939, pp. 396f.

[2] J. C. Eccles, *The Neurophysiological Basis of Mind*, 1953, pp. 278–79.

These unexpected facts have an instructive bearing on the relation between brain and mind. In my view the mistake of Descartes lay, not so much in his dualistic theory of a non-spatial, non-material mind, distinct from a material brain in physical space, but in his proposal to make the two interact at some central pontifical point in the solid spatial mass that makes up the brain. Here " it sits and governs ". His account thus treated the mind as the sole imperial arbiter of a totalitarian state, whereas what we now know of the organization of the brain itself suggests rather that it forms a vast society of living units which must be ruled on democratic principles.

(ii) But secondly, the activity exerted by the ' will ' must itself be an orderly activity. It can only function as an effective modifier of neuronal activity if its own activity exerts a patterned ' field of influence ', to borrow Eccles' phrase. In that case, as he says, the neural output that ensues will be determined by three factors: " (1) the micro-structure of the neural net and its functional properties "; " (2) the afferent input " (including presumably feedback); and " (3) the postulated ' field of extraneous influence ' " exerted by the agency we loosely call the mind or will.[2]

This then is a plausible hypothesis offered to psychology by contemporary neurology. It is avowedly no more than a tentative speculation. But it keeps within the limits prescribed by contemporary physics. The interesting feature is that, instead of the more familiar notion that the mind functions by detecting the activity of the brain, the brain functions by detecting the activity of the mind—and, under certain conditions, by failing to detect or so failing to transmit its influences. And, since the processes involved can best be treated by statistical techniques, the hypothesis becomes of special interest to the statistical psychologist.[3]

---

[1] As Eccles points out, this is consistent with the clinical evidence that " mind-brain liaison is not interrupted by the excision of a large part of the cerebral cortex or when a large part is occupied by intense activity that excludes liaison, i.e., by a convulsion of alpha-wave activity ".

[2] *Op. cit.*, p. 277.

[3] An extended corollary to this ' detector hypothesis ' might conceivably offer an explanation for some of the anomalous results reported in experiments on telepathy and the like. Let me add, however, that it throws no light on the special Cartesian problem raised by Dr. Smythies and others (*J. Psychic. Research*, XL, p. 395), namely, whether or not the mind itself is extended in space. I would only point out that the fact that certain mental *contents* are spatially extended does not (as he appears to maintain) prove that the *mind* is spatially extended. But, if we accept a relational theory of space itself the question ceases to have much importance. For certain purposes the factorist is of course ready to treat any event (e.g., the fact that n persons simultaneously perceive what they take to be the same object) as specifiable by a number of coordinates (here $3n$) and therefore in the mathematical sense a ' point ' in a multidimensional ' space '. But this is an analytic ' space ', not a concrete space in the sense intended by Dr. Levy and Dr. Smythies.

THE GENERAL AESTHETIC FACTOR. I

To the Editor, The British Journal of Statistical Psychology

DEAR SIR,—The correspondence on 'The General Aesthetic Factor' (this *Journal*, XIII, 87–92) is interesting for what it implicitly raises as much as for the explicit questions put by Mr. Paterson or the statistical evidence offered by Miss Williams and Sir Cyril Burt. I say this because I feel that in his letter Mr. Paterson has not pressed the strong case which he really has, and because the evidence offered is, in part at least, irrelevant to the problem.

Mr. Paterson's question concerns the existence of a General Aesthetic Factor, and how it is to be described if found. Remembering that we cannot get out of a statistical analysis more than we put into it, it behoves us to ask whether the original response-measures, which are used to generate correlations for the analyses, have any relevance to a theory of aesthetic 'appreciation' as opposed to one of 'preference' : or in other words, in what sense can the existence of consistent modes of hedonic evaluation, by a culturally homogeneous group of people with minimum intelligence not far short of average, be relevant to demonstrating that aesthetic appreciation is the perception (and presumably the valuing) of sense data in relational form ?

There is a minor issue of definition which has first to be settled—whether evaluative feeling-responses in other modes than vision and hearing can be aesthetic. It has, for instance, been maintained by some aestheticians that smell and taste preferences cannot be aesthetic. If we accept this view (on the ground that peculiarities in the phenomenology of taste and smell supposedly prevent the detached non-instrumental character of aesthetic evaluation from occurring), then either we do not demand that the factor analyst shall work with test-batteries covering *all* modalities, or if he insists on demonstrating factors involving these modalities, even when general intelligence and motivation are partialed out, then we could object that what he has found was not aesthetic by definition. For the purpose of argument then we shall accept as aesthetic some responses in all modalities, even including such recherché pleasures as having one's back scratched. The common denominator of the aesthetic tests cited in the evidence lies in the material being accepted by both experimenter and subject as generally desired for aesthetic reasons within the common culture of the two.

The problem is then one of interpretation ; and in aesthetics this presumably means either (*a*) that one is testing the theories of some aesthetician where these theories can be translated into operational language, which is more specific than " some people will express a preference for certain definable attributes of objects, and these preferences are themselves better on a metacriterion[1], than other preferences " (which seems to be the most one can get out of a great deal of aesthetic theory) or (*b*) that one is testing a theory based on psychological findings in other fields, such as perception or attitude studies, which can be generalized to aesthetics.

In the correspondence Professor Burt has drawn a parallel with Roger Fry's observations on the peculiar quality of the 'mental attitude' in aesthetic activity. Whether he has confirmed Fry's theory would seem to depend as much on what we accept as measuring operations, when we are investigating common qualities of mental attitudes in a diversity of aesthetic activities, as on the formal similarity between Burt's findings and Fry's assertions[2]. But, if we may be wise thirty years after Fry, a test of different acts of aesthetic evaluation in different contexts with very similar material (stylistically and iconographically) would seem more germane than Burt's evidence ; and we might then discover that the statistically insignificant residuals under the different methods of testing (this *Journal*, footnote 1, p. 91) would be somewhat greater. Incidentally, when comparing the results of aesthetic experiments, the use of the expression 'different methods of testing' is surely question-begging if interpreted as a statement with implications for aesthetic theory, and not simply as the language of an educational psychologist ; for it implies that there are 'right' responses and differences in 'ability' or 'performance' at tests which in some way reflect the capacity of the individual. Such a statement is not admissible in the context of subjectivist aesthetics, but belongs rather to the implicit absolutism which runs through Professor Burt's (1933) theorizing and his interpretation of his general factor[1].

---

[1] Used in the sense of a criterion to evaluate criteria, *vide* Carnap.

[2] The formal similarity is rather tenuous. Fry is writing of 'a certain positive response', which he 'conveniently labels . . . the specifically aesthetic state of mind'—surely a phenomenologically-defined common denominator of aesthetic experience which we should want to identify (if we identified it with a factor) with the first factor from a battery of tests in which the response-measures were content-analyses of introspective reports having far greater complexity than a mere

Oversimplifying somewhat in the interests of brevity we may say that our problem concerns the existence of a general aesthetic factor based primarily on hedonic preference-sorting, and not on judgements of aesthetic merit (a distinction investigated by Morris in *Varieties of Human Value*, 1956), such a factor can be accounted for in two ways : either (*a*) as evidence that aesthetic value exists and can be correctly perceived by an ideal critic to whom we all approximate in varying degrees—the extreme of absolutism— or in a weaker form, that all aesthetic evaluation will tend to take one of a small number of forms in a person exposed to certain material over sufficient time (because of some inherent aesthetic ability which is sufficiently similar in all persons) : or (*b*) at the opposite extreme it might be accounted for by a complete subjectivism in which all value-responses are correct at the time they are expressed, and subjects may tend to agree because of conditioning, group conformity, exposure to selected material, or something like those hedonic shifts which are discussed by Flugel under the heading of 'L'Appétit vient en Mangeant' (*Brit. J. Psychol.*, XXXVIII, 1948, pp. 171–190).

At one extreme the group factor is interpretable as the expression of that aesthetic perception of which we are all capable in some degree ; at the other extreme it is the artefact of a common culture imposing a conformity of responses on a population of subjects who will, within physiological limits, value any properties, direct or relational, or any object exhibiting such properties, for reasons which are a mixture of the aesthetic and non-aesthetic. It has been suggested[2] that both theories in their extreme form are statements about the use of language, and hence empirically untestable. Professor Burt has not indicated how his statistical demonstration could be part of an *experimentum crucis* between an absolutist and a subjectivist aesthetic theory ; and, until he does so, Mr. Paterson's question remains unanswered. It is, of course, probable that his question is not really one for psychology at all, but for theoretical aesthetics, and that this is disguised by being re-phrased in terms of a *post hoc* interpretation of statistical results.

The contrast between the two explanations of conformity in aesthetic judgements—that of 'culture' and that of 'innate ability'—has some similarity to the nature-nurture dispute in the assessment of intelligence ; and probably this has misled Mr. Paterson into putting the dichotomy in a rather misleading way, and Miss Williams into suggesting

---

preference. A choice situation is necessary and sufficient to elicit preferences, but neither necessary nor sufficient to identify that with which Fry was concerned. Burt has quoted it without regard to context. Operationally he has not shown that his tests differentiate between aesthetic choice and non-aesthetic choice except in terms of the material used, a distinction which Fry is anxious to make (*Transformations*, 1926, p. 2, third paragraph).

[1] In *How the Mind Works* (p. 307) we read: "the Venus of Milo would remain more lovely than the Queen Victoria statue in the Mall, the Taj Mahal than the Albert Memorial, though every man and woman in the world were killed by a passing comet's gas ". Apparently not only aesthetic value, but relative values, are absolute in this theory, which is perhaps derived from G. E. Moore's speculations on the fate of a Greek vase buried in the sands of Carthage for all time.

[2] Elton, *Aesthetics and Language* (1954), or Carver, *Aesthetics and the Problem of Meaning* (1952).

that Miss Conway's tests on twins will probably settle the issue. I suggest they will not, because the problem is not so simple as the ordinary nature-nurture dispute ; the measures in the aesthetic tests will have to differentiate between (i) hedonic responses, (ii) responses recognizing formal properties as present and rating them on quasi-aesthetic scales, and (iii) descriptive measures imposed on the material by the experimenters. These appear from the literature to be a hotch-potch of physical judgements and value judgements (' classical ', ' impressionistic ', ' colourful ', and the like). The absolutist theory, if accepted, would permit us to say that the subject whose responses identifying works of art as good are most typical (or heavily loaded with the general factor) will also be the subject whose personal preferences are most closely in agreement with what he identifies as good. The subjectivist theory, on the other hand, maintains that any relation between hedonic responses and judgements that ' art is good (by accepted canons) ' must be culture-loaded, and that the internal consistency of a judge with his own hedonic criteria should be logically independent of any conformity of his preferences with social norms, it would follow that it was also empirically independent, provided we can partial out the effects of previous exposure to material, or of differences in intelligence, or of any other variables likely to induce conformity for extra-aesthetic reasons.

If Mr. Paterson wishes to give vocational guidance on the basis of scores in aesthetic tests, what values of his own is he bringing to the interpretation of the general aesthetic factor ? If popular and hence commercially successful art is practised by persons whose loadings are different from (say) symphony orchestra musicians, would he select subjects whose artistic interests apparently outweigh their talents, by applying the principle of maximizing individual satisfaction, and would he then direct them into commercial art, or would he maximize potential aesthetic value for the community as a whole and then increase the number of unemployed artists whose repertoire is serious but whose execution borders on the inept ? The general aesthetic factor loadings of some of Washburn's subjects (who strongly approved the works of the Pre-Raphaelite Brethren) was doubtless high ; yet I would be reluctant to employ any of them as art gallery directors if they still held the same opinions.

Perhaps one of the protagonists of the ' general aesthetic factor ' as a test measure will defend the ephemeral nature of the test material used in experimental aesthetics, and tell us where ' tachiste art ' will be valued on the classical-romantic bipolar factor, and whether this bipolar would have had the same name if the test material had been composed exclusively of works by 18th century Romantics or 20th century Surrealists. To sum up my objections, although the so-called general aesthetic factor has been shown to exist, either as a fact or as an artefact, nevertheless its status within the context of aesthetic theory is ambiguous, and its validity as a selection criterion is heavily dependent on the culture-loaded and ephemeral material on which it is based. It follows that any psychologist who proposes to use it should always state quite clearly whether he is doing so in order to serve the interests of the applicant in the labour market, or to help in satisfying the man's artistic aspirations, or finally, to preserve and diffuse the aesthetic preferences of the tester himself.

ROBERT A. M. GREGSON.

THE GENERAL AESTHETIC FACTOR (Reply)

Mr. Gregson's difficulty seems largely to arise from a confusion between two kinds of general factor—that obtained by correlating tests and that obtained by correlating persons. Under certain conditions the two may be indirectly related : where the investigation has been planned so that both methods can be applied to the same data, the factor *measurements* obtained by the former are the factor *saturations* obtained by the latter, and vice versa. But here we have to do with two distinct sets of data.

Mr. Paterson's inquiry was concerned with an alleged general factor underlying aesthetic appreciation in all the different fields which it is possible to study. Such a factor can only be confirmed or negated by correlating *tests* for different types of aesthetic appreciation. At first sight it might seem that Mr. Gregson's questions relate to the precise nature of this general factor. But his main argument, if we understand him rightly, turns on the suggestion that the responses to the tests described by Miss Williams and Burt merely express the testee's personal tastes or 'hedonic preferences'; if so, he argues, they could not claim " any relevance to a theory of aesthetic appreciation". But, so far as this suggestion can be dealt with by statistical analysis, it calls for a very different type of approach and indeed a very different type of research from those on which Miss Williams was engaged. Such a question must be settled not by correlating tests, but by correlating *persons*, e.g., by comparing the different orders in which the individuals tested arrange the test-items. Even then the results need to be supplemented by introspections secured from the subjects who take part in the research. Correlations of this type (together with introspections) were given in the book which Mr. Gregson subsequently quotes (*How the Mind Works*, pp. 280f.). But they have little to do with Mr. Paterson's problem.

In some measure the issue which Mr. Gregson raises is, as he himself implies, a matter of definition, and therefore to a large extent merely verbal. Are we justified in applying the term 'aesthetic' to the general factors we have found ? In point of fact the word was originally adopted as a convenient label. Those who are engaged in the practical tasks which Mr. Gregson mentions—educational and vocational guidance—continually ask the psychologist whether he can supply any convenient tests for ability in drawing, painting, designing, and in the various branches of music. The tests described by Dewar, Williams, and Burt had been constructed for these and similar purposes; and the particular fact of psychology to which we now chiefly wished to draw attention was the existence of a *general* ability underlying skilled execution and judgement in each of these different fields. We needed a term which would be applicable to all these branches of activity. 'Artistic' (which is sometimes used) tends to suggest visual art to the exclusion of musical and poetic, and so we chose the word 'aesthetic'.

In the chapter on 'The Psychology of Art' which Mr. Gregson quotes, Burt definitely stated that *many* of the responses to the tests there employed (and indeed to many other tests) had little to do with aesthetic feeling in the stricter sense : " the majority of people ", it was said, " offer not aesthetic judgements but personal judgements ". That, however, does not necessarily imply that the tests are not tests of aesthetic appreciation, but merely that many persons, either from lack of endowment or perhaps from lack of training, do not exhibit any marked degree of aesthetic appreciation.

Is there then any evidence that those who score high marks on the tests may be validly described as possessing a general[1] capacity for aesthetic appreciation? Mr. Gregson asks if " the protagonists of the general aesthetic factor as a test measure will defend the ephemeral

[1] By way of validation the results of the separate tests were correlated with teachers' estimates for the *specific* type of ability with which each was concerned. But we suspect that Mr. Gregson would not accept the ordinary teacher as a sound judge even of specific aesthetic qualities. However as the third paragraph of his letter shows, this was not the issue he wished to raise. Both he and Burt (in the passage quoted) are concerned with the nature of aesthetic value in general, regardless of its manifestation in any particular medium.

nature of the material used"; and he seems to suppose that the material was restricted to nearly contemporary works of Western artists and so "culture-loaded". That, however, is scarcely correct. We do not claim to be "protagonists of the factor *as a test measure*" (a phrase used by neither of us); but we may perhaps point out that the visual material used in the tests included reproductions of paintings, statuary, architecture, carving, etc., from a wide variety of epochs and countries. They included paintings by Leonardo, Raphael, Michael Angelo, and Botticelli; such works as Praxiteles' Hermes, the Esquiline Venus, the Venus of Milo, and statuary by Donatello, Michael Angelo, and Henry Moore; and such buildings as the Parthenon and the Taj Mahal. Does Mr. Gregson consider these to be "merely ephemeral"? Moreover, in the whole series of investigations, including the earlier tests carried out over the wireless by Miss Margaret Bulley and Burt and the later investigations carried out by Miss Cancardas on men and women of widely different nationality, the replies obtained came from representatives of almost every available type of culture.

The astonishing thing was that practically all the correlations were positive. This implies the existence of a 'general factor for persons'. As we reported at the time, "though differences in taste and judgement undoubtedly exist, they are far smaller than the heated disputes and discussions on the subject would lead one to suppose". Now when the correlations between the several orders are positive, it is both legitimate and possible to derive a single order, based on a weighted average of the lot, to represent the 'general factor.' We then naturally inquire what kind of individuals—out of all those who took part in our experiment—have produced orders which agree most closely with this ideal or general order? The answer is that they are with few exceptions leading connoisseurs, or eminent performers in the particular field of art with which we are dealing; in other words, they are artists and art critics, musicians and musical critics, poets and literary critics, whom nearly every unbiased person would acknowledge to be each an aesthetic expert in his own particular field. It is this result, as Burt argued in his earlier articles, that justifies us in regarding the general factor as an 'aesthetic' factor.

Mr. Gregson asks whether in our researches we were "testing the theories of some aesthetician . . . or . . . a theory based on psychological findings in some other fields, . . . generalized to aesthetics". Our own answer is—"Neither. We were testing a theory suggested by a finding in separate fields of aesthetics—namely the existence of general factors for artistic (i.e., visual) appreciation, for musical appreciation and for literary appreciation respectively—a finding which we thought might be generalized to the whole field of aesthetics."

Mr. Gregson however supposes we are testing the theory of certain aestheticians who have favoured what he terms an absolutist theory of aesthetic value; and he complains that we have not "indicated how (our) statistical demonstration could be part of an *experimentum crucis* between an absolutist and a subjectivist aesthetic theory". Now in the reports of our statistical investigations there is nothing whatever to suggest that we were attempting to decide between two alternative theories of aesthetic *value*. Our problem was psychological not philosophical; our aim was to decide between two alternative theories of aesthetic *ability*—namely (i) the theory of a general ability entering into all fields coupled with specialized aptitudes peculiar to particular fields, and (ii) the theory of specific faculties for particular fields, with no general ability covering all the different fields—the theory adopted, implicitly if not explicitly, by most of the psychological writers who had previously touched upon the problem.

Now it is true that in the later book to which Mr. Gregson refers Burt certainly discussed the apprehension of value in its various forms—aesthetic, moral, and scientific (popularly identified with 'beauty', 'goodness', and 'truth'), and we both explicitly favoured an absolutist rather than a subjectivist view. The undeniable psychological findings which have commonly been cited in support of a subjectivist view appeared also in Burt's investigations (*op. cit.*, p. 288). But we argued that these merely showed that *some* of the judgements of nearly all of us and very possibly *all* the judgements of some of us were subjective, and it did not necessarily exclude the existence of an objective component in addition to the subjective.

As Mr. Gregson implies, the controversy over the objective or subjective nature of values is commonly supposed to be purely philosophical. But in our view there is no sharp dividing line between speculative philosophy and empirical science ; and psychology, even more than the other branches of natural science, often has evidence to offer on what were traditionally treated as problems of epistemology or metaphysics. Our own arguments for the objectivity of value were based on a relational theory of perception. The counter-arguments for the subjectivity of value are essentially the same as the familiar arguments for the subjectivity of our other perceptions, e.g., our apprehension of material objects in three-dimensional space. The latter depend on a confusion between two senses of the word perception—i.e., between what we perceive and the process of perceiving it. The latter is undoubtedly subjective, i.e., mental. It involves a relation between a conscious subject and the thing that he cognizes. But this does not imply that the thing cognized must also be mental. And the same reply will hold in the case of the values that we cognize. The fact that a round baize-covered card-table appears to one person oval and to another circular, to one green and to another grey does not prove that the table itself is purely subjective. And the fact that one design seems beautiful to this person, banal to another, and repulsive to a third does not prove that aesthetic value is itself subjective. Equally of course in neither case does our reply prove the opposite view. The objectivity of value, like the objectivity of material things, remains only a hypothesis that can claim the highest probability because of its explanatory power. Thus, to take almost the only argument involving statistical comparisons, the fact that the similarities between the rank-orders submitted by people representing such widely different cultural backgrounds are far greater than the differences seems to favour an objectivist rather than a subjectivist hypothesis. But the arguments we brought forward on *this* particular issue were not statistical, but observational or introspective. We maintained that our experiments demonstrated that what is commonly called aesthetic appreciation is essentially an ability to appreciate " shapes, colours, sounds, and even emotions *in certain relations* ". This not only suggested that the appreciation must be general, i.e., independent of the medium in which the relations were embodied (visible, audible, kinaesthetic, etc.), but also seemed to imply an objective existence for those relations. The use of the word ' beauty ' to describe aesthetic value is perhaps liable to be misleading, especially since, in these modern days, the word has come to be associated with the various cheap and sentimental forms of prettiness. What aesthetic intuition really detects, so we have argued, is the evidence of order in a world that is only partly ordered. But to justify this view at length would lead us far beyond the limits to which the present journal is confined.

E. D. WILLIAMS
C. BURT

BOOK REVIEWS

Psychological Twin Research. By TORSTEN HUSÉN. Stockholm: Almquist & Wiksell, 1959. Pp. 153. Sw. Kr. 18.

Dr. Husén's study of adult twins is not only the most recent but also one of the most thorough; and, in view of the renewed interest in mental genetics among British psychologists and educationists, it seems well worth while to examine his conclusions in some detail, and compare them with the results obtained in this country in similar studies carried out on children of school age. The data on which his investigations are based consists of the medical, educational and psychological information obtained for the twins discovered among the young men of 20 called up for National Service in Sweden between the years 1948 and 1951. The original aim, so he tells us, was to study in a large and representative sample " the hereditary-environmental conditions with special reference to intelligence test scores and school marks ". But at an early stage of the inquiry he found it necessary to start by examining, far more closely than previous investigators had done, " the methodological problems of twin research "; and this forms the main subject of the present publication.

The material collected related to nearly 3,000 individual twins. But the analysis here reported is restricted to two main samples: first, a group of 900 pairs, who constitute the basis of the main investigation; and secondly a ' control sample ' consisting of 50 pairs, living in the Stockholm area, who were subjected to a more intensive study in order to check the validity of certain of the diagnostic methods employed. The tests used were mainly verbal, and were applied to groups. They consisted of (i) Synonyms, (ii) Concept Discrimination, (iii) Directions (substituted after 1949 for the original Number Series), and (iv) Matrix. The statistical analysis is unusually thorough.

In general the results thus obtained with adult twins fully confirm the conclusions drawn from studies of children in the London surveys. To begin with, it is noted that in both the physical and mental development the average level of twins falls below that of the general population. In height and weight the difference was about 0·20 S.D.; and, as with the London twins, the inferiority was most marked in the case of monozygotic pairs. With the intelligence tests the average of the twins was 4 I.Q. points (about 0·25 S.D.) below that of the rest of the population. Dr. Husén reports that " there were nearly twice as many twins as single-borns with test scores below I.Q. 75 ". Similarly in London schools there was a decidedly greater proportion of twins in special schools for the mental deficient than in the ordinary elementary schools. The differences Dr. Husén attributes chiefly to disturbances during the pre-natal period and at birth. This also agrees with the evidence as to causal factors obtained from the case-histories in our own investigations.[1]

For height and weight the intra-class correlation amounted to 0·89 and 0·81 with the monozygotic pairs and 0·59 and 0·56 with the dizygotic group. The London figures were 0·95 and 0·91 for monozygotic pairs and 0·47 and 0·58 for dizygotic.[2] Dr. Husén notes, as we had done, that height is affected more by ' heredity ' and less by environment than weight. With the intelligence tests the correlations for the monozygotic pairs was 0·90 and for the dizygotic 0·70. Our figures with group tests were 0·91 and 0·54 respectively.[3] The fact that his figure for the dizygotic twins is somewhat higher than those found in earlier studies he himself attributes to the fact that his batch was confined to one sex only: we should be inclined to attribute it partly to the fact that the persons tested were older and the tests employed more heavily weighted with verbal items. As Dr. Husén observes, "the total variance, which is made up of hereditary variance and environmental variance (apart from errors of measurement), is the same for both standing height and test scores. But," he continues, "we know from previous investigations that environmental influence of height is rather small; hence the coefficient 0·89 for height should be considered almost entirely as an expression of hereditary influences." However, scores obtained with verbal tests are bound to be influenced in some degree by the cultural environment; and consequently it is admittedly rather rash, on the basis of mere analogy and in the absence of further evidence, to draw the same conclusion in regard to the influence of hereditary factors on test scores as we have drawn in regard to height.

---

[1] Cf. C. Burt and M. Howard, this *Journal*, IX, 1956, pp. 123f. and refs. Mr. Sandon reports similar results (*ibid.*, XII, 1959, p. 135.)

[2] C. Burt, *Brit. J. Educ. Psychol.*, XXV, 1953, pp. 168; cf. J. Conway, this *Journal*, XI, 1958, pp. 187. (The figures for monozygotic twins here include both those reared together and those reared apart.)

[3] Cf. Burt and Conway, *loc. cit. sup.*

How then are we to proceed? "By comparing the intra-class correlation for a relatively large sample of monozygotics and dizygotics reared apart, with corresponding values for such twin pairs reared together, we should," he contends, "be able to tackle these problems more readily: in this way the concrete significance of environmental difference could be ascertained." Unfortunately, however, in his monozygotic group comparatively few partners had been separated for a year or more. In the undetermined group, we are told, "there were a considerable number of pairs reared apart during longer periods or even from birth, but in these cases it was apparently impossible to carry out a 'similarity diagnosis' or a more intensive study". Nevertheless, his results, so far as they go, seem in full agreement with the conclusions drawn from separated monozygotic twins in the British inquiry.

The correlations for school marks averaged 0·77 with the monozygotic twins, but only 0·52 with the dizygotic. In six cases the monozygotic twins were sent to schools differing in level by "two or more steps" (e.g., primary and lower secondary with leaving certificate), and with these the average difference in the intelligence tests amounted to 6·5 I.Q. points. This is what we might expect from the nature of the tests employed, and confirms the view that the 'intelligence' tests used were in fact appreciably influenced by schooling. Among the London cases there were four in which only one monozygotic twin had been allocated to a secondary school of the 'grammar school' type; and here the average difference in the scores for the *group* tests amounted to 4·5 I.Q. points. With individual tests the difference was statistically insignificant.

Of the other comparisons made by Dr. Husén one of the most interesting relates to their handwriting. It is often said that "identical twins write almost identical hands". Certainly when the similarities are determined by a standardized scale of handwriting (though not when determined by impressions of general similarity or by the assessment of specific characteristics) most investigators report a greater resemblance for monozygotic twins than for dizygotic. Thelen, for example, who used the matching procedure found with German twins (aged 14 or over) a coefficient of 0·82 for the monozygotic but only 0·57 for the dizygotic; however, with twins of school age the difference between the two groups was much smaller.[1] In a group of 48 pairs Dr. Husén found for the 25 monozygotic pairs a 'mean similarity rating' of 2·93 and for the 23 dizygotic pairs one of 2·73. With such small samples, however, the difference was non-significant. He concludes that "intra-pair similarity in handwriting does not differentiate between the two types to anything like the same extent as intelligence test-scores, school marks, or anthropometric measurements".

One of the most interesting features of the whole research is the study of the various clues available for determining whether or not a given pair of twins is monozygotic or dizygotic. Hitherto the most instructive investigation of this methodological problem has been that of Essen-Möller.[1] To obtain a criterion he selected two distinct groups: (i) those twins where the records had reported only one chorion at birth, and treated this as a safe indication of monozygosity; (ii) those where either the sex or the blood group or both was different, and treated this as a safe indication of dizygosity. He then calculated the relative number of cases in which a classification based on various readily accessible clues agreed or disagreed with this basic classification. This enables him to express the value of each clue in terms of a probability coefficient. He then derives a formula for estimating the accuracy of a diagnosis based on all the available and independent clues, viz.

---

[1] E. Thelen, 'Zuordnungsversuche an Schriftproben von Zwillingen', *Zeitschr. f. Psychol.*, CXLVII, 1939, pp. 215f.

[2] Cf. E. Essen-Möller, 'Empirische Ahnlichkeitsdiagnose bei Zwillingen', *Hereditas*, XXVII, (1941), pp. 1f.

$$p=\frac{1}{1+\left(\frac{D_1}{M_1}\cdot\frac{D_2}{M_2}\cdots\frac{D_i}{M_i}\cdots\frac{D_n}{M_n}\right)P},$$

where p (the proportion of correct determinations) indicates the accuracy of a polysymptomatic diagnosis based on n symptoms, M_i and D_i the proportions of monozygotic and dizygotic twins showing the ith characteristic, and P = the ratio of dizygotic to monozygotic pairs (usually about 5 to 3 in the case of same sex twins). The structure of the formula becomes intuitively intelligible if we consider cases in which there is only one symptom: we then have $M/(M+D\times P)$; and this = 1 when $D=0$.

Dr. Husén's ' control investigation ', as he calls it, was based on 50 pairs—26 being monozygotic and 24 dizygotic. These were subjected to a detailed examination (including a standardized interview) lasting for about two hours. His methods to a large extent follow those of Essen-Möller. However, even when it was possible to discover the hospital in which the twins were born, the information available seldom indicated whether or not there was only one chorion at birth; and consequently his criterion for monozygosity was a ' final diagnosis ' based on all the measurements and general information available. Tests for four uncorrelated blood group systems (ABO, MNS, Rh, and P) were taken by the staff of the Institute of Genetics, and analyses were obtained from the National Chemico-Legal Laboratories. It appeared that, if a pair belongs to the same blood group in all four systems, the probability that they were dizygotic was only 7 per cent and that they were monozygotic 90 per cent, the MNS and Rh systems being the most informative, since in these the dizygotic pairs were dissimilar in 12 out of 22 cases. A diagnosis based on 10 monozygotic characteristics gave probabilities of 76 and 18 per cent. Among the other characteristics the frontal hairline, the colour of the hair, and the complexion was nearly always the same in the monozygotes (the exceptions being less than 9 per cent). They were also the same, however, in about half the dizygotes. Of the other characteristics nose, lips and teeth were strongly suggestive, but shape of ears (contrary to the common view) proved far less reliable. The study of the portraits is also instructive. With full face photographs the 22 monozygotes were given an average dissimilarity rating of less than 20 and all except 4 of the dizygotes a rating of more than 20. One obvious clue, however, Dr. Husén has apparently omitted, which other investigators have regarded as highly suggestive, namely, the comparison of fingerprints.

According to the replies given by the monozygotes 28 per cent had sometimes been mistaken for each other by their parents, 19 per cent by their siblings, 85 per cent by their friends, and 89 per cent by their teachers. Dr. Husén therefore believes that " we can depend to a considerable extent " on inferences drawn from such replies. With London children we found that the twin's own intimate playfellows rarely made mistakes, and that mistakes were quite exceptional in a normal home circle. At school and elsewhere girls were confused more easily than boys—chiefly because among the girls identical twins tend to adopt the same dress and hair-style, and secretly enjoyed the confusion so caused. Apart from the recorded details of the birth, we found the best single clue was the general impression of the experienced examiner: when further evidence was subsequently available it appeared that, if a pair was classed as ' identical ', the diagnosis hardly ever proved wrong; but a few errors were discovered where the pair had been classified as ' dizygotic '. With this Dr. Husén's results concur. About 10 per cent of the monozygotes were, he states, wrongly judged to be dizygotes; and he thinks that similar mistakes may have occurred in earlier inquiries so that many monozygotic pairs have been wrongly included in the dizygotic group. "Hence," he says, "the part played by environmental variance has been exaggerated."

In nearly all attempts to disentangle environmental influences from genetics there is one troublesome circumstance which is too often disregarded, namely, the correlation or 'interaction' between the two. Most investigators have too often assumed that the quality of the environment and the amount of environmental variance are the same for all members of the same family, whether monozygotic twins, dizygotic twins, or ordinary siblings. Common observation, however, suggests—and systematic comparisons[1] have fully confirmed—that environmental influences will show greater similarity in the case of twins than with other siblings even when of the same sex. 'Identical' twins are notoriously more closely attached; they go about together; they cultivate same interests; and they share in the same enterprises. Often there may be a tacit agreement that one of the two shall act as spokesman, but this does not prevent their opinions and decisions from being joint. These are facts familiar to every school teacher from personal observation of his own pupils. Dr. Husén has taken up this point in studying his group of adult twins. His method was to address a formal questionnaire, suggested by the findings of earlier investigators, to the twins themselves. He found that the 'attachment score' (based on a quantitative analysis of the replies) "was of the same order of magnitude in both the DZ and the MZ groups", and "in both groups the tendency towards attachment is lower than during their childhood. . . . What differentiates the two twin groups most is the striving to be dissimilar manifested by the DZ twins."

Dr. Husén's whole research is one that might well be emulated by investigators in this country. Primarily, as he has emphasized throughout, it is a study in methodology; and, as such, will be invaluable to those planning similar inquiries. But, quite apart from the details, there are several broader lessons that the genetical psychologist can readily draw.

(i) First of all, the analysis of data already available, such as test-scores or school marks, though useful in a pilot-inquiry and suggestive as a starting-point, is inadequate for purposes of conclusive research. The ordinary group tests of 'intelligence' are too deeply affected by educational and other environmental conditions to throw much light on genetic differences. Far more trustworthy methods of assessment are needed, e.g., individual tests, depending as little as possible on the cultural background, planned and applied specifically for the research in hand, preferably by two independent investigators working in cooperation, and checked whenever possible, independently by the observations of those who already know the examinees, such as their parents and teachers.

(ii) As Dr. Husén himself has argued, by far the most convincing method of distinguishing environmental effects from genetic is to compare cases in which the 'identical twins' have been brought up in homes of a widely different type. The prejudice against separating twins is not so strong in the case of new-born infants, and, as Burt has pointed out, there are far more cases of separated twins than is commonly realized. With the cooperation of officials, and under a guarantee that confidential information will be duly respected, it is by no means difficult as time goes on to collect and examine a large number of cases.

(iii) It is important that accurate records should be preserved of the details relating to deliveries of twins so that future investigators are not confronted with the difficulties Dr. Husén and others so frequently experience in deciding which are monozygotic and which are not.

(iv) Dr. Husén is undoubtedly right in giving the highest priority to the study of methodology—i.e., of the most appropriate experimental designs and the effective statistical techniques. He himself (again quite rightly) starts with Fisher's intra-class correlation coefficient rather than with Pearson's product moment coefficient. But it may be remembered Fisher himself went on to express the intra-class correlation in terms of the variances due (*a*) the hypothetical contributory factors, i.e., environment and heredity, and (*b*) to the observable components, i.e., the variance within families and the variance between families. This leads to a method of estimation which Burt has called 'the factorial analysis of variance', developed on the assumption of a multifactorial Mendelian inheritance. How far Fisher's techniques need to be modified or amplified in the light of those since developed by Mather and Sewall Wright are further problems in methodology calling for practical trial.

---

[1] Cf. H. von Bracken, 'Mutual Intimacy in Twins: Types of Social Structure in Identical and Fraternal Twins', *Character and Personality*, II, 1934, pp. 293f. For typical anecdotes illustrating 'common observation' see Galton, *Inquiries into Human Faculty*, pp. 157f.

(v) From Dr. Husén's work it also seems clear that researches on mental genetics can be carried out far more effectively and far more readily with children than with adults. Children are easier to test; they can be interviewed more readily and more fully and the relevant information is more readily accessible. Comparative data for adults is, as Dr. Husén's investigations demonstrate, highly desirable; but this, it would seem, could best be secured by following up cases that have been first studied during the school period.

(v) Owing to the supreme importance of the questions involved the inquiries should be carried out on an official basis. Large numbers of subjects and large masses of accurate information are essential. And this of itself requires time, and money, and trained personnel. The unofficial and amateurish efforts of isolated research-students is likely to add little to existing knowledge. But to secure adequate cooperation it will be necessary to arouse public interest in these branches of research as well as in those more spectacular branches of scientific research which have hitherto chiefly attracted official and popular attention.

J. CONWAY

Chance, Skill, and Luck. By JOHN COHEN. Harmondsworth: Penguin Books, 1960. Pp. 201. 3*s*. 6*d*.

This is an attractive book with an unattractive title. As he explains in his prefatory chapter, Professor Cohen's real aim is to analyse those modes of behaviour which involve that superlatively human activity we call "guessing", or, in other words, conduct that entails "taking risks". His treatment, he tells us, will be psychological—that is partly experimental, partly statistical, partly introspective, but not behaviouristic in the sense of being essentially neurological. The neurological model for man, he tells us, was set up by psychologists not by neurologists. "Its fundamental weakness is its impersonal character"; and risk-taking is an eminently personal affair.

He begins with a brief historical introduction. This leads him to trace the parallel between such ancient social institutions as casting lots, trying by ordeal, and resorting to oracles, and other forms of divination, and the modern individual customs of trusting luck, looking for omens, and reading inferences from horoscopes as offered in the newspaper. The scientific study of the subject is of comparatively recent origin. "Un problème posé à un austère janseniste par un homme du monde", says Poisson, "a été l'origine du calcul des probabilités." The "problem" arose out of certain bets placed by an inveterate but thoughtful gambler. But the "austere jansenist" considered only the rational, not the emotional or personal factors, and concentrated solely on the mathematics. Indeed, as Burt has remarked, "from that day to this the *psychology* of gambling and betting has never yet been properly studied": the gambler, he says, as assessed by those who know him, is "of an extraverted rather than introverted type and of a sanguine or at least superficially euphoric type : half his time he lives in a world of wish fulfilment". Temperamentally he shows many striking resemblances to the typical delinquent, and "though most gamblers may not be criminals, nearly every criminal is in outlook an habitual gambler". "While the pessimistic introvert trembles before a risk of 1 in 200, the optimistic extravert—and that includes many criminals—will ignore a risk of 1 in 3 or 4 if it runs counter to his wishes, but treat it as a certainty if it would favour what he desires." Similarly Professor Cohen notes that a chance of 1 in 7,000 for winning a prize is regarded as well worth accepting by the hardened gambler ; but the same chance of dying as a result of crossing the road would be considered negligible by nearly everyone. There is one further point of similarity between gambling and crime. Since the war, gambling, like delinquency, has considerably increased. " In 1910 the turnover of the gambling industry amounted to about £100,000,000 ; by 1938 it had risen to approximately £380,000,000 ; by 1947 to more than £700,000,000 ; and it is still increasing."[1]

Professor Cohen is interested in the subject from the standpoint of general psychology rather than of individual psychology; and logically no doubt the former aspect should be investigated first. He views it as a problem in probability; and his mathematical procedures are largely based on those with which psychological students will already be familiar from their textbooks on statistics. But he distinguishes two types of probability—objective probability (which the statistician investigates) and subjective probability, or, as he now prefers to say, psychological probability (which the psychologist investigates). I myself, however, would venture to suggest that the real subject should rather be called ' the intensity of conviction ' : " the variable to be assessed and studied is not the strength of a certain type of probability, but the strength of certain beliefs ". Professor Mace in his Foreword gives a neat illustration. Just before the last general election a celebrated statistician told an eminent man of business that his analysis of the probabilities strongly indicated a win for the Conservatives. After the election they met again, and the business-man reported : " I made a handsome profit by buying Steels: and you? " "Alas," said the statistician, " I lacked the courage of my conviction." " The odd conclusion follows," adds Professor Mace " namely that the statistician did not really *believe* what as a statistician he knew to be true—that the *probability* of a Conservative victory was very high." In short, as he puts it, the ultimate effect of these investigations should be to " restore to psychology the concept of belief, which has almost disappeared from the textbook ".

CHARLOTTE BANKS

Numerical Methods of Curve Fitting. By P. G. GUEST. Cambridge: University Press, 1961. Pp. xiv+422. 80*s*.

The author's aim is to give a comprehensive account of methods of combining or reducing sets of observations. Part I deals with observations of a single variable. Under this head he discusses the general theory of frequency- and probability-distributions, the postulates leading to the arithmetic mean (least squares, minimum variance, and maximum likelihood), the properties of the normal curve and of the commoner discrete distributions, and the relevant statistical tests. Factorial design and analysis of variance are deliberately omitted. We then turn to cases where the observations are concerned with two variables. Part II describes the fitting of straight lines and the theory of regression equations. Special attention is given to the problems arising when the observations are not equally spaced, and when they are not of equal weight, and finally when the independent variable is itself subject to error. Part III, which occupies nearly two-thirds of the volume, deals with the fitting of polynomials and of non-linear functions which are not polynomial. This leads to a discussion of the general regression equation and problems involving functional relations between several variables. Free use is made of matrix notation to express and deduce the more complicated results, but a knowledge of only the simplest matrix operations is assumed.

Throughout the book full derivations are given for the formulae used. Each method is illustrated, not only by a worked example, but also by a complete calculating scheme, so that the beginner can proceed along well tried paths. The schemes presented for the more complex procedures are designed primarily for a desk-computer. Tables needed for the practical calculations are printed at the end of each chapter. A couple of alternative procedures are given for calculating orthogonal polynomials where high speed automatic computers are used, but most of the schemes are suitable for either method of computation. A full bibliography of over 200 numbers is appended; and the whole book is admirably produced.

---

[1] C. Burt, *The Contributions of Psychology to Social Problems* (Oxford University Press), pp. 61f.; cf. *id.*, *The Young Delinquent*, p. 152. Professor Cohen's figures are somewhat smaller; but he observes that the estimates can only be very approximate. In any case he fully confirms the remarkable increase. In his preface Professor Mace also suggests that the psychological study of gambling may have an important bearing on the problem of crime.

The examples, while highly interesting in themselves, are restricted to physics ; and the types of observations discussed are those most frequently encountered in the physical sciences. However, the author hopes that the book will also be useful to those engaged in the biological sciences. Since many of the techniques presented here find no place in the usual statistical textbooks, it should also prove an invaluable work of reference for research workers in psychology.

CYRIL BURT

Attitude Scaling. Edited by J. DOWNHAM. London: The Oakwood Press, 1960. Pp. iv + 76. 15*s*.

In this country psychological researches on ' attitudes ', as they are now called, started nearly fifty years ago with studies of the attitudes of school children towards their school, their teachers, and most frequently of all towards the subjects of their curriculum. The concept of ' attitudes ' was based largely on McDougall's theory of ' interests ' or ' sentiments '. These are all alternative names for much the same thing, namely, " those complex conative dispositions which are concerned largely with valuation rather than cognition, and which determine in a more or less consistent fashion the subject's behaviour." The earliest studies treated the problem as essentially qualitative. Later the concept was somewhat simplified, and attempts at quantitative scaling were introduced. As a rule, the procedures used were based either on simple ranking or on one of the traditional ' psycho-physical methods '. Finally, a fresh attempt was made to deal with the more complex aspects of the subject by the application of factor analysis. With the founding of the National Institute of Industrial Psychology this type of research spread into the vocational and commercial fields; and about 1930 the publication of several highly instructive American studies, first by Thurstone and his co-workers, and then by Likert, Cantrill, and many others, aroused a lively interest in the statistical techniques.

The present volume contains a series of papers presented at the Second Conference held by the Market Research Society, which is also responsible for the publication. The nine contributors are British investigators, all of whom are, or have been, engaged in research on personal or social attitudes. Most of the discussion is concerned with methodology—the measurement of attitudes by formal interviewing and the choice of statistical techniques for analysing attitudes and making reliable predictions. Norman Squirrell, Market Research Officer for a well known firm, writes an introduction to the series. Of the papers that follow, one of the most useful—' A General Introduction to Attitude Scaling Techniques '—is by J. B. Handyside of the National Institute of Industrial Psychology. D. L. Lamberth, Operations Research Manager at Messrs. J. Lyons and Co., briefly discusses the place of attitude assessments in what is termed ' operations research '; Joseph Trenaman, who at the time of the Conference was Further Education Officer at the B.B.C., gives an admirable account of Guttman's ' scalogram ' technique; R. P. Kelvin, of University College London, describes how rational, irrational, and unconscious factors can influence the attitudes of consumers; and J. D. Downing describes an instructive experiment planned to study the attitudes of potential customers towards two different brands of the same commodity.

The chief interest of the whole collection springs from the fact that it arises out of the co-operative efforts of scientific investigators who represent widely different fields of application and approach them from widely different standpoints. It should serve as a valuable introduction for younger British workers, and might well be followed by a further symposium dealing not so much with methods as with typical results.

I. E. ROBERTS

Introduction to Econometrics. By Oskar Lange. London: Pergamon Press, 1959. Pp. 384. 35*s.*

This book is based on a series of lectures given by Professor Lange in the University of Warsaw and translated from the Polish by Eugene Lepa. The synthetic word 'econometrics'—modelled on Galton's term 'biometrics'—was coined by Ragnar Frisch, the Norwegian statistician and economist, in 1927. And, as biometrics soon became largely independent of biology, so econometrics seems to be developing a similar independence of traditional economics. According to the present author it has three main aims: first, to determine economic trends and so to forecast future situations, notably the business cycles which he regards as a special feature of capitalist societies; secondly, to investigate the problems of demand and supply, that is, of 'market research' in the broadest sense; thirdly, to lay the bases for planning or programming. Each of these fields is discussed in turn in the present volume—the first most briefly, and the last in considerable detail. "Economics", it has been said, " is itself a branch of applied psychology, since the phenomena with which it deals are the result of human discrimination, valuation, and motivation ". Hence, just as the economist has much to learn from psychology, so the psychologist could learn a good deal from the methods and results of the economist as set forth in this volume.

" The first econometric research ", we are told, was that which led Pareto to formulate the law of income distribution that goes by his name. In its cumulative forms the law is expressed by the simple equation, $N = C/x^a$, where N is the number of persons whose income exceeds x units, C is a constant, and the exponent a—a so-called 'coefficient of elasticity'—measures the inequality of the incomes (Pareto suggests it varies about 1·5 or rather less). Or, differentiating, we can write $y = aC/x^{a+1}$, where y denotes the proportionate number of persons having an income of $x \pm \frac{1}{2}dx$ units. Professor Lange takes two examples of income distribution, one for Britain in 1911, and a second for the U.S.A. in 1919; and shows that, particularly in the case of the British data, the fit is quite good. American economists have maintained that the equation really expresses " a natural law of the distribution of human abilities "; and Professor Lange himself points out that a similar J-shaped distribution results from the effects of a series of examinations taken during successive years of study: the largest proportion pass the first examination, a smaller proportion the next, and so on. However, he believes that, as applied to incomes, the curve is really the effect of the unnatural type of successive screening that applies chiefly in capitalist countries. The present distribution of incomes among employees in Poland, he finds, does not fit Pareto's formula, but is rather " reminiscent of the right-hand half of the normal curve ". And he quotes other investigators who report that, with relatively homogeneous social groups, the distributions nearly always approximate to the normal, though the curve is usually truncated asymmetrically at either end. He is, however, unable to offer any convincing explanation of the curves obtained or of their divergences and odd anomalies; and he appeals, quite rightly, for further research.

Now, as Burt long ago pointed out in discussing this and other equations, the distribution of abilities is approximately normal, as we should expect from the multifactorial theory of inheritance. But a very similar equation has been derived by Marshall and his followers to express the relations of demand and supply when what he calls the 'elasticity' is constant ($d = C/p^a$, where d denotes demand and p price). This suggests that motivation, as well as ability, is involved. Pareto's law, according to Burt, expresses not the distribution of ability, but the distribution of output, and output depends on *both* psychological variables ('Ability and Income', *Brit. J. Educ. Psychol.*, XIII, 1943, pp. 93f.). Their influence on one another is expressed not by adding the influences, but by multiplying them (cf. Table IV, *loc. cit.*, p. 96). If the motivational factor were nearly the same for every individual in the sample, the distribution of output would be approximately normal; otherwise it would approach the asymmetrical and J-shaped distributions such as are described by Professor Lange.

By far the longest chapter, nearly half the volume, is devoted to the theory of 'programming'—a term that is used to describe " economic planning when the planning is not on the scale of an entire country ". As Professor Lange points out, the theory of the subject developed in the U.S.A. during World War II in connection with problems of supply for the armed forces and with the organization of the forces themselves. It is a topic to which he himself has made important contributions, particularly in that branch of the subject known as 'input-output analysis'. The problems emerging he treats very largely by the aid of matrix algebra, deducing 'characteristic equations' similar to those involved in 'factor analysis by weighted summation'. This part of the book is naturally of greater interest to economists than to psychologists. Nevertheless, in their formal aspects the questions that arise are very similar to many which the psychologist encounters. And accordingly psychologists as well as economists should be very grateful for the enterprise which has resulted in the translation and publication of this highly illuminating volume.

I. E. Roberts

Social Mobility in Industrial Society. By S. M. Lipset and R. Bendix. London: Heinemann. California: University of California Press, 1959. Pp. xxii + 309. 30*s*.

Dr. Lipset and Dr. Bendix deal with three main problems: first, social mobility in general as a characteristic of human communities; secondly, social mobility in one particular group, treated as a representative community, viz., the East Bay section of metropolitan San Francisco; and thirdly the causes, dimensions, and consequences of social mobility. The more general discussions are based on a wide survey of the relevant literature.

The authors' opening chapter leads up to what they rightly describe as " a startling conclusion ", namely, that the degree of social mobility seems to be much the same in nearly all industrial societies at nearly every period of time. As they point out, the opinion hitherto most frequently voiced by American writers is very different from this: the United States, it is assumed, form an 'open' society with considerable mobility, while most European countries (England, France, Germany, etc.) have societies that are 'closed', so that children of workers are more or less forced to remain in the social position of their parents, and children from the upper classes remain in the upper classes. A parallel view obtains over here. As regards this country there was in the past, so most British educational and social psychologists apparently believe, hardly anything approaching social mobility until quite late in the nineteenth century; but during the past 50 years or so, owing to improved standards of culture among the working classes, and the great increase in free places at secondary or grammar schools,and above all to the steady weakening of the barriers separating one class from another, it has rapidly become very much easier for children of the so-called lower classes to rise into the higher. Perfect freedom, it is admitted, has not yet been attained; but the abolition of the 11-plus examination, the general introduction of comprehensive schools, and other measures of educational and social reform, will in the end " virtually banish class differences". And the same writers commonly point to America as " a model country whose educational system we might do well to emulate " and to certain other backward countries in Europe as even more immobile than ourselves.

What then is the evidence of the facts so far as they are now available? Mobility, as our authors point out, is of two main types—intra-generational and inter-generational. Researches on intra-generational mobility seem to be comparatively few. In a British survey carried out in 1949 by Geoffrey Thomas it was found that of persons whose first job had been of a manual type as many as 21 per cent had, by the time of the survey, risen to an occupation of a non-manual type; and similarly of those who started with a non-manual job 23 per cent had dropped to a humbler manual employment. Similar investigations have been carried out in America and Japan, with similar results: the chief difference—by no means a large one—is that in these other countries it is commoner for people to rise to higher jobs than to fall to lower jobs.

Studies of inter-generational mobility are more numerous. Accordingly for purposes of their main comparison Lipset and Bendix define mobility as the proportionate degree to which sons of manual workers rise into non-manual classes or vice versa; and their evidence is based on quantitative data collected in various countries since the end of the last war. In view of the recent discussion of this and kindred problems in the previous issues of this *Journal* it is perhaps worth while bringing together the actual figures given in their tables and diagrams.

TABLE I. DEGREE OF SOCIAL MOBILITY IN VARIOUS INDUSTRIALIZED COUNTRIES

	Upward Mobility	*Downward Mobility*	*Total Mobility*
Urban Only			
United States A	35	25	30
United States B	31	34	
France	35	18	27
Germany A	27	38	31
Germany B	30	20	
Germany C	26	28	
Sweden	31	24	29
Switzerland	44	13	23
Japan	33	21	27
Urban and Rural			
Great Britain	20	49	29
Denmark	22	44	31
Italy	8	34	16

Note.—In the first nine surveys the figures in the first two columns give the percentages of the total number of sons of manual workers who have risen to non-manual occupations and of sons of non-manual workers who have taken up manual occupations. In the last three the distinction is between occupations of high and low prestige, which is not always identical with the distinction between non-manual and manual work. In particular, in Denmark and Italy, and perhaps even in Great Britain, the inclusion of farmers, etc., may very possibly have exaggerated the amount of downward mobility. Accordingly an attempt is made to derive a single overall index of vertical mobility by adding the raw figures on which the first two columns are based and then reducing the result once more to a percentage.

It is unfortunate that the methods of classification are not always based on the same criteria. But one might expect this to exaggerate rather than diminish the differences. Nevertheless, the data seem to give a fairly reliable indication of the fluidity of urban populations. Except perhaps for Switzerland and Italy the amount of social mobility (as defined by the authors) is much the same in all the countries studied as in the U.S.A.

The main conclusion is not entirely new. In his book on *Social Mobility* (1927) Pitirim Sorokim undertook a detailed examination of the data then available, and came unhesitatingly to the conclusion that, contrary to the prevalent view, there was no greater mobility in the United States than in the industrialized countries of Europe. Similarly Professor Glass, in the investigation discussed in an earlier issue of this *Journal* (X, pp. 10f.), compared the amount of mobility at various periods from the end of the last century and during the earlier decades of the present century, and could find little sign of an increase in spite of the popular impression to the contrary (*Social Mobility in Britain*, 1954).

If this were a complete and reliable account of the matter, it would provide a strong refutation of the views held by those educational and social psychologists to whom I have already referred: for it would seem to imply that, however much we improve the education of the underprivileged and however much we may succeed in abolishing class distinctions, the amount of upward mobility would remain unchanged. Nevertheless, the rather crude method of assessing mobility on which the investigations summarized in Table I are for the most part based is in my view both incomplete and unconvincing. The division of the population into manual and non-manual is only one of the many class distinctions that it is possible to draw. Moreover, in the majority of the investigations cited the term ' non-manual ' includes relatively humble white collar jobs such as those of routine clerks as well as the upper professional classes. But when the working class boy is seeking to rise to a higher social level his aims soar much further than that of a non-manual worker on an office stool. In most cases what he hopes to do is first to win admission to a grammar school, and then eventually to achieve either university education and a place among the professional classes or else its equivalent in the world of commerce or industrial management.

Now when we consider mobility into and out of the professional classes the story is quite different from that which the two American authors describe. Take, for example, the social origins of higher civil servants in this country (those above the rank of Assistant Secretary): in 1929 only 6·6 per cent had fathers in the Civil Service; in 1939 the proportion had risen to 9·5, and in 1950 to 15·6.[1] Much the same chance can be traced in the history of the medical, legal, and teaching professions. Moreover quite large variations are found when instead of comparing times we compare places. In Britain, France, Germany, and Sweden a smaller proportion of civil servants and professional men are drawn from the manual classes than is the case in America, but a greater proportion among the members of the legislature and the cabinets. This may be partly related to the fact that in America far more youths attend universities than in Europe, while labour parties as such have a remoter political role. But even here we must be careful not to generalize too rashly. If we turn from the higher professional strata to the analogous strata in commerce and industry, the variations are far less striking. As long ago as 1912 an inquiry undertaken in the British cotton industry by Professor Sidney Chapman and Fred Marquis (now Lord Woolton) showed that over two-thirds of the owners, directors, and managers had begun their careers as manual workers or routine clerks.[2] And a series of investigations carried out in various countries indicates that the social backgrounds of American and European ' business élites ' is very similar. Mrs. Floud too has found that today the proportion of male teachers drawn from the manual classes is actually a good deal smaller than it was before the war ; and studies of members of legislatures and cabinets indicate that in Europe far more politicians are drawn from the lower social classes than is the case in America.

One other restriction affects the figures I have quoted. They relate solely to social mobility among males. During the last seventy years or so there has been a vast extension of the number of non-manual occupations open to women—commercial, clerical, and even professional; and different countries vary enormously in the extent to which women are able to take advantage of these new opportunities.

But the question that will chiefly interest readers of this journal will be the extent to which social mobility is related to the psychological characteristics of the individuals and particularly of the children. According to Dr. Lipset and his colleague, "in a bureaucratized industrial society the main source of occupational achievement is educational achievement". Certainly this is very true of the United States, where, we are told, " a college education has become an almost necessary requisite for any position of responsibility "; in Britain a university degree is considered essential only for entrance to one of the professions. Now educational achievement is dependent on the individual's intelligence. Hence we may suppose that the limitations of an individual's inborn capacity are likely to limit his powers of rising to a high position. On the other hand, the mere possession of a high measure of innate ability does not of itself guarantee a high occupational level: this is amply shown by follow-up studies of junior county scholars in this country and of gifted children in America.[1] Health, motivation, social advantages, all play an appreciable part.

---

[1] R. K. Kelsall, *Higher Civil Servants in Britain from* 1870 *to the Present Day*. 1950. Pp. 150f.

[2] ' The Recruiting of Employing Classes from Ranks of Wage Earners ', *J. Roy. Statist. Soc.*, LXXII, 1912, pp. 293f.

Dr. Lipset and Dr. Bendix conclude that social mobility may be regarded as the outcome of two main factors: first, "the structure of opportunities to which the individuals are exposed"; secondly, " their capacity to take advantage of these opportunities "; to these I would add a third, the family background. " Capacity ", as they understand it, " consists of motivation or drive as much as it consists of intelligence. But even in the U.S.A., it appears, educational achievement is to some extent dependent on the financial resources of the family; and a good deal of the drive is imparted by the family. Much is effected by " directly urging the children to achieve ". Moreover, among the working classes, " the higher the education of the parents, the more they tend to instil a desire for higher education in the children "; and in America, it must be remembered, a number of fathers in the working classes went on from high school to college; often the mothers themselves have come from a non-manual class. Above all, we are told, members of small families are more likely to rise than members of large families. For this there are several reasons. To begin with, as the authors point out, the parents who deliberately restrict the size of their families are likely to be more intelligent, better educated, more self-controlled and more ambitious to do their best for their children. At the same time the child in the small family finds it easier to pick up adult ideas, adult information, and mature ways of speech. And these differences are found not only between the manual and non-manual classes, but also between different families within the manual class.

The final conclusion that the authors draw is one which, as they put it, " has travelled some distance from the viewpoint of the nineteenth century ". " Our findings," they say, " suggest that, at the basis of our social structure, the problem is not merely one of 'natural endowment thwarted by social inequality', as was previously assumed ". In a country which is " second to none in its concern with mobility " it seems reasonable to demand that if, after all, individuals differ as widely in their intellectual endowment as many geneticists are beginning to believe, then each should have an occupation suited to his abilities. Dr. Lipset and Dr. Bendix, however, believe that their review reveals that the apparent misfits and the dissatisfactions that result are due not so much to traditional prejudice or to the natural jealousy which in a democratic country the less fortunately endowed may feel towards those whose talents secure a quick advancement; they are due rather to inadequate motivation, particularly to " a lack of interest in educational and occupational achievement ". And here the authors believe lies one of the most urgent problems for future research.

M. HOWARD

Readings in Social Psychology. Edited by E. E. MACCOBY, T. M. NEWCOMB and E. L. HARTLEY. London: Methuen. New York: Henry Holt, 1959. Pp. xii + 674. 55*s*.

This is the third edition of a volume which first appeared just after the second world war. Since McDougall published his *Introduction* half a century ago, the interest in social psychology has developed at an ever-increasing pace. Quite half the material that appeared in the second edition of these ' readings ' had not appeared when the first edition was published; and nearly a third of the present edition has been selected from the voluminous mass of literature that has been issued during the last five years.

The book makes no claim to cover the whole field; it is intended, we are told, merely as " an illustrative selection of empirical studies and approaches which may supplement more systematic presentations ". The extracts are almost entirely confined to American writers. Readers in this country will miss references to problems which have formed major subjects for research among British psychologists—delinquency and crime, the effects of poverty and slum conditions, social mobility, and above all the influence of genetic factors. " Personality characteristics ", says one of the contributors, " develop as a function of the environment "; and that is the standpoint adopted by most of the contributors. The idea that part at least of the manifold differences in intelligence or temperament exhibited by different individuals and by different social groups might conceivably be the result of differences in innate make-up is tacitly dismissed as if it were an out-of-date superstition. On the other hand, the British reader will find that the book provides a most valuable survey of the main fields of recent American work.

---

[1] See C. Burt, ' The Gifted Child ', *Yearbook of Education*, 1961.

Of the newer topics which found no representation in the earlier edition the most important is that of ' perception '. Two of the thirteen main classifications into which the whole book is subdivided are now concerned with this topic; and Jerome S. Bruner contributes a highly instructive article on ' social psychology and perception ', specially prepared for the present volume. Those who are not already acquainted with this new line of interest will be surprised to find that so ancient a topic bulks so large. The explanation is, not merely that older psychologists were more interested in the physical determinants of perception than in its social determinants, but that the word is here used in a very broad sense: it includes, not only what old-fashioned writers called ' apperception ', but almost every type of ' interpretation '—including, for instance, the use of such formulae as $s=gt^2/2$ to describe the course of falling bodies. The emphasis on a term like ' perception ', which seems to savour more of the earlier introspective psychology of this country than of the rigorous behaviourism of American writers, is perhaps an indication that the social psychologist has at length discovered that for his purposes a rigorous behaviourism is not enough.

Other new topics to which increased attention is devoted are those of public opinion and group dynamics. One far-reaching innovation, however, is not represented here by any addition to the content: during recent years researches in social psychology have tended more and more to employ rigorous experimental design and advanced statistical techniques. The editors comment on the change; but they feel that the students for whom they are writing will not have the technical background needed for understanding these more scientific procedures and the arguments and conclusions based on their use. They have, they say, " attempted to eliminate high-level statistics, such as analysis of variance and factor analysis "—techniques which originated in Britain and which have therefore formed favourite tools with British investigators. At the same time it should perhaps be acknowledged that this strong preference for quantitative techniques has sometimes closed our own eyes to many wider problems and many alternative modes of approach. Hence for many British readers such a book as this may perhaps come as a welcome eye-opener.

J. M. DAVIS

The Future of Man. By P. B. MEDAWAR. London: Methuen, 1960. Pp. 128. 10*s.* 6*d.*

This book contains the Reith lectures broadcast during the winter of 1959. An appendix of notes has been added with detailed evidence and up-to-date references for the views advanced in the text; and this in many ways forms the most valuable part of the volume. The somewhat ambitious title was apparently chosen not by the lecturer but by the B.B.C. Professor Medawar begins with an introductory talk stressing the fallibility of all predictions about ' the future of man ', and goes on to discuss the limits that are set to future improvement by biological conditions. This is followed by a lucid and authoritative account of the ' genetic system of man '—a section which every student of psychology, whether or not he is familiar with the biological approach, should certainly ponder.

To readers of this journal the lectures of greatest interest will be those concerned with mental inheritance and the genetics of intelligence. In this field, as Professor Medawar reminds us, confusion has reigned until quite recently, because most psychological writers have clung to the theory of blending; and the few who have rejected it have assumed that the simple form of Mendelian inheritance—so-called ' segregative ' or ' unifactorial inheritance '—is the sole or at least by far the commonest type. In point of fact variations in quantitative characteristics such as intelligence and stature depend in the main on multifactorial, or, as Professor Medawar prefers to call it, on ' metrical ' or ' polygenic ' inheritance. In spite of the far-reaching influence of environment, he is fully agreed that " differences in intelligence are strongly inherited ".

In the appendix Professor Medawar refers to the ' naïvetés ' of certain psychologists, and criticizes what he calls " Burt's belief that the aptitudes revealed by intelligence tests are inborn and virtually uninfluenced by upbringing and environment ". I find it difficult to discover how any words of mine could have suggested this interpretation. Nevertheless two or three of my recent critics have quoted this passage as showing (in the words of one of them) how " a leading biologist has rejected the conclusions drawn by Howard, Burt, and others in their attempts to apply a multifactorial hypothesis to intelligence ". As a matter of fact in the notes to his lecture on ' Intelligence and Fertility ' Professor Medawar cites a pamphlet of my own with the same title; and, so far as I can see, the views which I there expressed are very similar to those that he himself reaches in his admirable discussion of the problem. I myself referred to the criticisms commonly passed on " psychologists who naïvely assume that the tests measure pure native ability "; and I stated that, to the best of my knowledge, no competent psychologist had ever entertained such a notion. All admit " the I.Q.'s furnished by intelligence tests are affected in *some* degree by schooling and other environmental influences ". The crucial question therefore is—in what degree ? The answer, I maintained, must usually depend on the nature and number of the tests employed and the type of population tested: with group tests, mainly verbal and applied once only, the influence will be rather large; if the crude results are checked and corrected by individual tests and repeated examination it will be much smaller. In general, I suggested, we might assume that " one half of the variance at the very least is attributable to ' nature ' as distinct from ' nurture '." My main evidence was based, not only on the correlations obtained between ordinary children and their parents, but also on intensive studies of " children brought up in residential institutions " and of " identical twins, particularly those reared separately from early years ".

Each type of evidence has been severely criticized both by behaviourists and by environmentalists. Professor Medawar, however, accepts all three lines of argument as valid. " The study of correlations in the population generally ", he concludes, " and particularly of their values in foster-children and identical twins reared apart, suggest that not less than half the observed variation is an inborn variation ". On the final issue therefore biology and psychology seem in complete accord.

He then goes on to discuss the possibility of a gradual decline in national intelligence. " The argument for a decline ", he says, " is based on the belief that differences of intelligence are under the control of a multitude of genes, and it is assumed that their contributions to intelligence are additive." There is, he continues, no reason to doubt that, over the normal range, inborn differences of intelligence result from polygenic transmission. This provides an emphatic reply to those who have condemned the multifactorial theory. The additive nature of the various contributions may, he thinks, be rather more open to question; but, after reviewing the evidence, he accepts this assumption also. And his final conclusion is that, " if innately unintelligent people continue to have larger families, then, with some qualifications, we can infer that the average level of intelligence will decline ", though, he contends, " the decline will be a slow one ".

His last lecture returns once again to his main theme—" the future of man ". Here he describes " the origin of a new, non-genetical system of evolution based on properties of the brain "—namely, its ability to learn and hand on what it has learned. In a sense this new mode of evolution might be termed Lamarckian: socially, though not genetically, we do inherit acquired characters. Much of our future evolution, so far from being fatally fixed, is in our own hands; it is, or can be, " the subject of human decisions and of acts of mind ". Nevertheless, this does not mean that all inequality between individuals and groups can be abolished by a series of social and political reforms. " Mendelian heredity ", which applies, so far as we know, to the whole of genetic inheritance, itself involves a " conservation of inborn diversity ". " It works in such a way that the proportion of individuals having one genetic make-up or another can be assumed to remain pretty constant from generation to generation, unless something special happens to prevent it." Thus " a large part of human fitness is vested in a mechanism which provides for a high degree of genetic inequality and inborn diversity, which makes sure that there are plenty of different kinds of human beings, and sets a limit to any theoretical fancies we may care to indulge about the perfectibility of man ". We can thus improve upon nature; but the very possibility of doing so " depends on our continuing to explore nature ". CYRIL BURT

Principles of Comparative Psychology. By R. H. WATERS, D. A. RETHLINGSHAFER and W. E. CALDWELL. New York: McGraw-Hill Book Company, 1960. Pp. 453. 62*s*.

This volume is the joint work of seventeen specialists, each of whom deals with his own particular branch. As the preface emphasizes, the subject, unlike that of many other works with a similar title, is not animal psychology but comparative psychology in the literal sense, that is, " a comparative study of the psychological similarities and differences between species". After a brief introduction defining in greater detail the problems to be examined, the book opens with a discussion of classification. At present, as the authors rightly point out, there is nothing like an acceptable taxonomy of behavioural characteristics. However, although they themselves use terms familiar to the statistical psychologist—'correlation', 'levels', 'dimensions showing quantitative variation', and the like—they seem to use them merely in a descriptive or metaphorical sense. They mention one tentative scheme—very similar to the 'hierarchical' classification of human characteristics favoured by British factorists—the scheme that begins with a twofold vertical classification into cognitive and motivational characteristics, and then introduces horizontal subdivisions for the sensory, perceptual, associative and other levels. But the suggestion appears to rest merely on an unconfirmed impressionistic basis. In this country, however, a more exact approach has been proposed by one or two psychologists, and welcomed by biologists like Julian Huxley, namely, a systematic application of the formal methods of factor analysis.

The book includes what the editors call " a new look at the old problem of innate behaviour ". Dr. King and Dr. Nichols, who contribute an admirable chapter on this subject, begin by quoting Asch's account of the unfortunate consequences that followed the type of physiological approach adopted by the behaviouristic school in the earlier decades of the present century. " The aim ", he says, was to build up a psychology " based predominantly on the study of lower organisms and to generalize the resulting principles to the human level "; and the outcome was, " a radical omission or devaluation of specifically human characteristics". Behaviour was reduced to two types—reflex activities on the one hand and activities that were supposed to be wholly learned on the other. The concepts of instinct and of innate tendencies or abilities were rejected or denied; and as a consequence, we are told, even in comparative psychology " the major research emphasis in this country " (the United States) " has been for the past twenty-five years on learning and motivation ". This they deplore; and they support their arguments for the new look by an excellent summary of recent work on innate behaviour carried out by Lashley, Lorenz, Tinbergen, and many other experimentalists.

Dr. Fuller deals with recent work on genetics and individual differences. He too insists on the value of a more intensive study of the innate components of animal behaviour.

He explains and illustrates the techniques available, and summarizes the conclusions already reached. "The inference from this array of experiments", he declares, "is that positive results are obtained rather easily when experiments are properly designed; and the conclusion that hereditary differences have widely ramifying effects on many aspects of behaviour is strongly supported."

Here and elsewhere stress is laid on the growing recognition of the value of statistical and quantitative procedures, and special attention is paid to the ingenious statistical theories of perception and learning recently put forward. Nearly all the chapters are illustrated by diagrams and photographs; and to each is appended an up-to-date list of references.

G. R. ROBINSON

Physics and Philosophy: The Revolution in Modern Science. By W. HEISENBERG. London: Allen & Unwin, 1959. Pp. 176. 15*s*.

A glance through the earlier volumes of this journal is sufficient to show how surprisingly indifferent most British psychologists seem to be towards the remarkable revolutions in scientific thinking that have been introduced by modern physicists during the past twenty-five years. And yet, as one or two of the papers contributed to the general discussion amply indicate, the discoveries and conclusions of relativity and of quantum theory in its second phase have a close bearing, not only on such concrete problems as the functioning of the brain and the mutation of genes, but also on the more abstract issues regarding the nature and function of consciousness in both its cognitive and its conative aspects. The study of quantum phenomena, we have been told, "has brought to light a number of non-classical properties and principles which characterize ultra-microscopic phenomena—indeterminacy, complementarity, the Pauli principle, the need for partial second quantization, and for abstract representation in higher complex spaces with a high degree of invariance, and finally the amazing pluralistic world of nucleons, high energies, transient particles, and 'strong' and 'weak' interaction—all flying in the face of the basic assumptions of all nineteenth century science".[1] Added to this there is the curious fact that the mathematical and statistical methods now used in quantum theory bear a singular resemblance to those employed in statistical psychology.

And, as if to clinch all this, we now have an illuminating survey of the whole field from the pen of the most eminent of living authorities on quantum physics—one who has himself been responsible for some of its most revolutionary principles and techniques. In an earlier volume, *The Physical Principles of Quantum Theory*, Dr. Heisenberg gave us a systematic survey of the chief experimental results and the main theoretical inferences; and here in his latest work his purpose is to consider "how these new conclusions permit or require man to think of himself in relation to his universe?" The substance of the book originally formed the Gifford Lectures delivered in 1955-6 at the University of St. Andrews. The whole is prefaced by a long but admirable introduction by Dr. F. S. C. Northrop, Professor of Philosophy at Yale.

Dr. Heisenberg begins with a brief history of his subject; and for this he goes back to the beginnings of the atomic theory, as propounded by the early Greek materialists in the 5th century B.C. He reviews the general development of physical science in successive stages —up to and during the Renaissance and then from Descartes and Newton down to Einstein. At each stage he brings out in clear relief the influence of current physical theories on the philosophy of the day and the bearing of current philosophy on the physical theories. Two chapters are devoted to stating and defending the Copenhagen interpretation of quantum phenomena—an interpretation which gave an entirely new turn to quantum theory. Dr. Heisenberg himself now tells us in his own vivid way how one day in 1926 discussions with Bohr at Copenhagen went on until late at night: "at the end I went alone for a walk in the park and asked myself again and again: ' Can nature be so absurd as it seems in these atomic experiments ? ' ". The solution that he reached is well known: " If ", he says, " one takes account of the limitations expressed by the ' uncertainty relation ', the contradictions disappear ". And he now goes on to argue that the ' uncertainty principle '—coupled with the surprising discovery that matter, so far from being conserved, can be either ' created out of energy ' or ' annihilated into energy '—really leads us back to the old Aristotelian doctrine; " the matter of Aristotle, which is mere 'potentiality', can be compared to our concept of energy, which gets into ' actuality ' by means of the ' form ', when the elementary particle is created ".

---

[1] C. Burt, 'Quantum Theory and the Principle of Indeterminacy', this *Journal*, XI, 1958, pp. 75–93; cf. *ibid.*, 31f., XII, pp. 153f., and XIII, pp. 185f.

To the psychologist the most interesting chapter is that on the " Relation of Quantum Theory to Natural Science ". I would humbly suggest that those who have followed the long controversies in this journal with the more thoroughgoing representatives of mechanism and behaviourism, like Dr. Taylor and Messrs. Clarke, Hammond, and Hansel, might fruitfully study Dr. Heisenberg's discussion of the problem. Dr. Heisenberg himself ruthlessly rejects the nineteenth century doctrine that science must postulate universal causation. However, Professor Northrop proposes a convenient compromise based on a verbal distinction: *all* causal systems, he suggests, need not necessarily be deterministic systems; but every deterministic system is necessarily causal. Thus he would maintain that, although in the physical sciences 'probable predictions' must now take the place of ' exact and certain predictions,' we can still call the process ' causal '.

When we turn from the physical to the biological sciences, we find, says Heisenberg, that " something has to be added to the laws of physics and chemistry before biological phenomena can be completely understood ". For example, living organisms " display a degree of stability which a complicated structure of many different types of molecules could not have on the basis of physical and chemical laws alone ".

The problems of psychology he illuminates by invoking Bohr's principle of complementarity. The consequence is, as Northrop points out, that the " Cartesian theory of a mental substance " and the dualism which this theory implies " come off very much better in this book than in any other contemporary work of the philosophy of physics ". Yet here too Heisenberg is inclined to revert to the older Aristotelian conception of the soul. " This ", he says, " as found for instance in the philosophy of Thomas Aquinas was more natural and less forced that the Cartesian concept of the ' res cogitans ' ". Those psychologists who maintain that it is " the duty of a scientific psychology to explain psychological phenomena in the terms of the physical behaviour of physical entities " should note that the physicists whose lives have been spent in studying " the physical behaviour of physical entities" are the first to reject all such maxims. For the physicist, so Heisenberg tells us, mind is a kind of mirror in which he sees the material world; and yet the modern science of psychology " seems to pay more attention to the neuron's mechanical properties than to its optical properties ". After all the mind is not merely a machine: it is an organism which cognizes.

The book concludes with a couple of chapters on ' language and reality ' and ' the role of modern physics in human thinking '. All our experiments, we are reminded, must begin and end in observable 'facts' such as form part of the phenomena in our everyday macroscopic world. It is characteristic of such facts that they can be described in terms of common concepts and ordinary language. But our common concepts and our ordinary language " cannot be applied to the structure of the atom "; that can only be discussed by means of mathematical symbols and mathematical arguments. Thus " in our experiments about atomic events we have to do with phenomena quite as real as any in daily life; but the hypothetical atoms and elementary particles are not so real: they form a world of potentialities or possibilities rather than part of the world of facts ".

There are two further differences between classical and contemporary physics which deserve the attention of the psychologist. Nineteenth century science assumed that the physical universe could be studied as an isolated system quite independent of the observing mind which studies it. Quantum theory and relativity have demonstrated that, since the observer himself forms part of the same system, his standpoint and his observations on the system must—at least in principle—always form parts of the total scientific description. There is even in the purest of physical investigations a psychological element which defies elimination. Secondly, it used to be maintained that science could and should be studied in complete independence of philosophy; but Heisenberg has little difficulty in demonstrating that physics is bound to introduce more physical and philosophical assumptions than the facts alone can warrant or supply. And the most fundamental of these are philosophical assumptions. They are, as the foregoing summary indicates, partly ontological—referring to the subject matter of physics in itself—and partly epistemological—arising out of the cognitive relation between the scientist as observer and the physical system that he is endeavouring to observe. Thus contemporary physics have brought about " a drastic revision of man's conception of the universe and of his own relation to it "; and the change that has thus taken place " is not simply a continuation or development of the previous picture; it involves a real break in the structure of modern science ".

The whole book is highly readable. It requires little or no knowledge of the technicalities of mathematics or physics; and it certainly deserves to be studied by every psychologist as part of the background of his scientific training.

W. L. Gregory

Abilities of Twins. By T. Huésn. *Scand. J. Psychol.*, I, 1960, pp. 125–35.

This article, which appears in a new journal, forms a valuable supplement to the monograph reviewed above. From the results of a " nation-wide programme of achievement tests " Dr. Husén has selected data relating to 634 pairs of twins. With every test their average was 1/4 to 1/3 S.D. below that of the single borns. For identical twins the intra-class correlation was 0·85 to 0·90; for non-identical twins of like sex 0·74, and for those of unlike sex 0·54. As Dr. Husén points out, the higher figures might be due to similarity in the environment. It is to be hoped that he may later be able to supply us with figures for twins reared apart.

The graphs and detailed figures for tests will be of special interest to workers in this field. In the main the results thus obtained with school children confirm the conclusions drawn in Husén's researches on adults. The distribution-curves, however, demonstrate that " the previous assumption that the inferiority was accounted for by a higher frequency of defectives proved invalid ".

Several correspondents have asked for a proof of the formula (p. 81) used for distinguishing identical twins. Professor Burt explains that the problem is a case of the ' combination of testimony '. " In a bag containing balls the proportion of white to black is p_0 to $1-p_0$. Let the relative frequency with which the testimonies of the first and second witnesses are correct be p_1 and p_2. A ball is drawn: if both say it is white, what is the probability that this is true?

" Of the 8 *a priori* hypotheses only two are relevant: (A) the ball is white and both witnesses are correct; (B) the ball is black and both are incorrect. The probability of A is $p_0 p_1 p_2$, of B $(1-p_0)\ (1-p_1)\ (1-p_2)$. Hence the *a posteriori* probability that the ball is white when both say it is will be

$$p=\frac{1}{1+\dfrac{(1-p_0)}{p_0}\cdot\dfrac{(1-p_1)}{p_1}\cdot\dfrac{(1-p_2)}{p_2}}.$$

Putting $(1-p_0)/p_0=\mathrm{P}$, this gives the equation used by Husén in his various inquiries."

J. Conway

OCCUPATIONAL PSYCHOLOGY

Editor : Alec Rodger

OCTOBER, 1960

VOLUME 34, No. 4

ood Assessment and Food Acceptance as a Psychological Theme. By R. Harper

The Development and Use of a Taste Panel Technique—A Review. By A. S. C. Ehrenberg and J. M. Shewan

s in the Measurement of Food Preferences by Triangular Tests. By R. A. M. Gregson

Absolute Judgements and Operative–Inspector Relations. By I. D. Brown

Semi-Skilled and Over Forty. By A. Heron and S. M. Chown

ge and the Timing of Movement. By K. F. M. Murrell and B. Forsaith

Book Reviews.

ther Books Received.

Annual Subscription **40** *shillings*

National Institute of Industrial Psychology, 14 Welbeck Street, London, W.1

2 NOV 1961

BRITISH JOURNAL OF STATISTICAL PSYCHOLOGY

Vol. XIV Part 1

May 1961

CONTENTS

		Page
Kaiser, H. F.	A Note on Guttman's Lower Bound for the Number of Common Factors	1
Burt, C.	Intelligence and Social Mobility . . .	3
Stone, M.	Subjective Discrimination as a Statistical Method .	25
Miller, C. R. Eyman, R. K., and Dingman, H. F.	Factor Analysis, Latent Structure Analysis and Mental Typology . . .	29
Burroughs, G. E. R., and Miller, H. W. L.	The Rotation of Principal Components . .	35
Notes and Correspondence		
	Factor Analysis and its Neurological Basis . .	53
	The General Aesthetic Factor	72
Book Reviews	. . . .	79

JOURNAL COMMITTEE

John Whitfield, *Editor*

Cyril Burt, *Assistant Editor*

and

D. E. Broadbent
B. Foss
M. Hamilton
A. N. Oppenheim
J. S. Small
J. Sutherland

Charlotte Banks, *Secretary*

Vol. XIV Part 2 November 1961

THE BRITISH JOURNAL OF STATISTICAL PSYCHOLOGY

EDITED BY
JOHN WHITFIELD

WITH THE ASSISTANCE OF
CYRIL BURT
AND THE FOLLOWING EDITORIAL BOARD

A. C. AITKEN
R. J. AUDLEY
M. S. BARTLETT
W. G. EMMETT
M. G. KENDALL
D. N. LAWLEY
E. S. PEARSON
E. A. PEEL
L. S. PENROSE
J. FRASER ROBERTS
A. RODGER
W. STEPHENSON
A. STUART
P. E. VERNON

Printed and Published by
TAYLOR & FRANCIS LTD.
RED LION COURT, FLEET STREET, LONDON, E.C.4

Price 20*s*. *per part* (*U.S.A.* $3.50) *Subscription* 30*s*. 6*d*. *per volume* (*U.S.A* $4.50)

PUBLICATIONS OF THE BRITISH PSYCHOLOGICAL SOCIETY

The British Journal of Statistical Psychology is issued by the British Psychological Society. The subscription price per volume is 30*s.* net for Members of the Society and 30*s.* 6*d.* (post free) for non-members. The subscription price in the U.S.A. is $4.50 (post free). Members of the Society should send their subscriptions to THE SECRETARY, British Psychological Society, Tavistock House South, Tavistock Square, London, W.C.1: non-members to Messrs. TAYLOR & FRANCIS, LTD., 18 Red Lion Court, Fleet Street, London, E.C.4. Until further notice the Journal will be issued in two parts each year, in May and November.

Papers for publication and books for review should be sent to the Assistant Editor, Sir Cyril Burt, 9 Elsworthy Road, Hampstead, London, N.W.3. Authors should retain copies of their manuscripts as neither the Society nor its Editors can undertake any responsibility for loss or delay in the case of unsolicited contributions. Authors will receive 50 copies of their articles free; extra copies may be purchased provided the order is given when the proofs are returned. Contributors should indicate whether they wish their reprints to be bound in covers. The printers will state current prices when sending out proofs.

The Society also issues quarterly *The British Journal of Psychology* and *The British Journal of Medical Psychology*: subscriptions, 60*s.* per annum for either journal, should be sent to the Cambridge University Press, Bentley House, Euston Road, London, N.W.1. The Society publishes, jointly with the Association of Teachers in Colleges and Departments of Education, *The British Journal of Educational Psychology*: for this the subscription, 30*s.* per annum, should be sent to Methuen & Co., Ltd., 36 Essex Street, London, W.C.2. Members of the Society receive the foregoing journals on special terms; enquiries should be addressed to the Secretary, British Psychological Society.

Papers for publication in *The British Journal of Psychology* should be sent to Dr. Boris Semeonoff, Department of Psychology, the University, Edinburgh; those for publication in *The British Journal of Medical Psychology* to Dr. Joseph Sandler, 96 Portland Place, London, W.1; those for publication in *The British Journal of Educational Psychology* to Professor P. E. Vernon, London University Institute of Education, Malet Street, London, W.C.1.

PREPARATION OF PAPERS

Contributors are asked to send their papers in a form which is ready for submission to the printer: that is to say, the arrangement of the material should follow that observed in previous publications of this *Journal*. Each article should be headed either with a series of numbered headings in italic, corresponding with the cross-headings of the several sections, or with a brief abstract: (the latter procedure is preferred). Each section should have a short cross-heading, in capitals; and the whole should conclude with a summary embodying the main conclusions reached. Special attention should be paid to such details as correct spelling, grammar, capitalization, numbering, etc. (particularly in the case of tables and references). In preparing manuscripts and in correcting proofs authors should conform, so far as possible, with the 'Recommendations to Authors' set out in *The Printing of Mathematics* by T. W. Chaundy, P. R. Barrett, and Charles Batey (Oxford University Press, 1954, ch. II, pp. 21–73), or to those given in the paper on 'Setting Mathematics' (Arthur Phillips, *Monotype Recorder*, XL, iv, 1956).

Owing to the present conditions for printing, the cost of setting up tables and algebraic formulae which require hand composition is very much greater than the cost of machine composition. Contributors are therefore advised to arrange their material so that it can be printed as economically as possible. Often numerical results and algebraic formulae can, with a little simplification, be readily adapted for purposes of machine composition. Manuscripts involving many tables and numerous equations can as a rule only be accepted if the authors are willing to pay for the additional cost. Contributors should note that the *Journal* is intended for the general psychological reader, not solely for those who specialize in statistics. Unusual technicalities should be either avoided or carefully explained, e.g. by reference to previous discussions of the same or kindred problems in this *Journal*.

In future authors will receive one complete slip-proof of their papers, but no page proof. They will be expected to pay the cost of all corrections for which they themselves are responsible.

Vol. XIV Part 2 The British Journal of Statistical Psychology November 1961

AN ASYMMETRIC APPROACH TO MULTIPLE-FACTOR ANALYSIS

By W. A. Gibson

Department of Defence, U.S.A.[1]

After a brief history of asymmetric approaches in factor analysis and related fields, there is derived a new asymmetric procedure that avoids communality estimation and other difficulties. This new solution splits a test battery 1 into sub-batteries x and y, each of which spans the same factor space as does a test battery 2. This permits finding the unknown transformation that converts any factorization of the between-battery correlations R_{12} into an orthogonal factor matrix reproducing both R_{12} and R_{xy}, the between-sub-battery correlations.

I. Introduction

The symmetry of the correlation matrix used in factor analysis is something for which we are often justly grateful; and there are well-known advantages in the symmetry of other matrices with which we deal both in that and in other related procedures. On the other hand, recent developments [1, 4] in the field of latent structure analysis have shown that an important gain may be secured by the deliberate *asymmetric* treatment of symmetric matrices, or rather from the concentration on asymmetric sections. The gain consists in the avoidance of certain elements with repeated subscripts that occur in the complete symmetric matrices, and this has a close bearing on the communality problem of factor analysis.

Already there are asymmetric approaches in factor analysis whose principal advantage is that of avoiding the need for communality estimation. One of these [5] is a solution for the proportional profiles problem introduced by Cattell [2, 3]; another the inter-battery method of Tucker [9]. The present paper will derive yet a further asymmetric procedure that is most closely related to that of Tucker. This paper indeed stems from an effort to overcome a weakness in Tucker's procedure that has been discussed in detail elsewhere [6, 7].

In the latent structure solutions just mentioned two symmetric empirical matrices are used—a matrix of pairwise joint frequencies and a matrix of third- or higher-order joint frequencies. Corresponding asymmetric subsections are drawn from them for the solution. The same thing is true for the asymmetric solution of the proportional profiles problem, where the two symmetric matrices from which asymmetric subsections are drawn are correlation matrices for the same variables but based on two different samples differentially selected on orthogonal factors. The Tucker approach strives to get along with but one empirical matrix and one asymmetric subsection therefrom. The result is that, in general, the solution does not give orthogonal factor matrices [6, 7]. A good way of summarizing the asymmetric approach to be derived here is to say that it attempts to make up for the use of but one empirical matrix by using two related but non-overlapping subsections therefrom.

---

[1] Opinions expressed are the author's, not those of the U.S. government.

II. Basic Development

It will be instructive to relate the formal development to the analysis of the fictitious example shown in Table I. This same example appeared elsewhere [6], with a correction on its analysis there being made in [7]. In Table I are shown correlations R among eight fictitious tests grouped into a battery 1 and a battery 2, with battery 1 further subdivided into sub-batteries x and y. The grouping is such that the sub-matrix R_{xy}, showing correlations between sub-batteries x and y, has the same rank and refers to the same factor space as does R_{12}, the sub-matrix of correlations between batteries 1 and 2. R_{xy} and R_{12} are blocked in and labelled in Table I. It is the requirement that R_{xy} and R_{12} must imply the very same factor space that is crucial here. Tucker's inter-battery approach uses only an R_{12}. An orthogonal factor matrix F_0 that completely reproduces R is shown at the right in Table I.

It is appropriate to introduce now the general topic of factorization of asymmetric (possibly non-square) matrices such as R_{12}. Here all that is meant by a factoring of an asymmetric matrix R_{12}, of rank r, is the finding of a pair of r-columned and rank r matrices A_1 and A_2, of proper length, such that

$$R_{12}=A_1A_2'. \tag{1}$$

For the present fictitious R_{12}, where $r=2$, a proper A_1 is simply the first two columns of R_{12}. Such an A_1 is blocked in and labelled in Table I. That A_1 has rank 2, and weighted sums of its columns will reproduce all columns of R_{12}. The necessary weights are, in fact, the various columns of A'_2, obtainable by standard solutions to systems of linear equations. In the present case, A'_2 can be seen by inspection to be simply

$$\begin{bmatrix} 1{\cdot}000 & 0{\cdot}000 & 1{\cdot}000 & 0{\cdot}000 \\ 0{\cdot}000 & 1{\cdot}000 & 0{\cdot}000 & 1{\cdot}000 \end{bmatrix}.$$

Other ways of factoring R_{12} will be discussed later.

Let us now turn to the more customary meaning of factoring and inquire how we might find a single orthogonal factor matrix F such that its various parts would reproduce both R_{xy} and R_{12} according to the fundamental factor theorem ([8], pp. 79-81). That is,

$$R_{xy}=F_xF'_y, \tag{2}$$

and

$$R_{12}=F_1F'_2. \tag{3}$$

Here the subscripts on the right indicate the portions of F corresponding to the sub-batteries and batteries that are involved. Now the similar form of eqns. (1)

TABLE I. INTERCORRELATIONS ($R=F_0F'_0$) AND ORTHOGONAL FACTOR MATRIX (F_0) FOR A FICTITIOUS TWO-FACTOR EXAMPLE

			$R=F_0F'_0$									F_0	
			Battery 1				Battery 2						
			Sub-Battery x		Sub-Battery y								
			1	3	2	4	5	7	6	8		A	B
Battery 1	Sub-Battery x	1	1·000	0·000	1·000	0·000	0·566	−0·707	0·566	−0·707	1	1·000	0·000
		3	0·000	0·640	0·000	0·640	0·453	0·566	0·453	0·566	3	0·000	0·800
					R_{xy}								
	Sub-Battery y	2	1·000	0·000	1·000	0·000	0·566	−0·707	0·566	−0·707	2	1·000	0·000
		4	0·000	0·640	0·000	0·640	0·453	0·566	0·453	0·566	4	0·000	0·800
							A_1		R_{12}				
Battery 2		5	0·566	0·453	0·566	0·453	0·640	0·000	0·640	0·000	5	0·566	0·566
		7	−0·707	0·566	−0·707	0·566	0·000	1·000	0·000	1·000	7	−0·707	0·707
		6	0·566	0·453	0·566	0·453	0·640	0·000	0·640	0·000	6	0·566	0·566
		8	−0·707	0·566	−0·707	0·566	0·000	1·000	0·000	1·000	8	−0·707	0·707

and (3) implies the existence of a non-singular linear transformation T such that

$$A_1T = F_1 \tag{4}$$

and

$$T^{-1}A'_2 = F'_2, \tag{5}$$

for substitution of eqns. (4) and (5) into eqn. (3) yields eqn. (1).

Thus the problem becomes one of finding the transformation T that converts the A_1 and A_2 satisfying eqn. (1) into the orthogonal factor matrices F_1 and F_2. For this purpose eqn. (4) may be split into two parts corresponding to sub-batteries x and y. Thus,

$$A_xT = F_x \tag{6}$$

and

$$A_yT = F_y. \tag{7}$$

Here A_x and A_y are the two parts of A_1 that are associated with sub-batteries x and y, respectively. Now substitution of eqns. (6) and (7) into eqn. (2) yields

$$R_{xy} = A_xTT'A'_y, \tag{8}$$

from which the least-squares expression for TT' is obtainable as

$$TT' = (A'_xA_x)^{-1}\, A'_xR_{xy}A_y(A'_yA_y)^{-1}. \tag{9}$$

The matrix TT' for the present example, computed according to eqn. (9), is shown at the top left in Table II. Below it is shown a triangular factorization, T, of TT', obtained by what is known, for other reasons, as diagonal factoring ([8], pp. 101–105). To the right of TT' and T in Table II is shown the matrix F_1, computed according to eqn. (4). Beside F_1 is shown F_2, computed by least-squares from eqn. (3) according to the formula

$$F'_2 = (F'_1F_1)^{-1}\, F'_1R_{12}. \tag{10}$$

F_1 and F_2 make up the orthogonal factor matrix F that we are seeking. Comparison of this F and the F_0 of Table I, or of their graphs, indicates that F is essentially a 45 degree rotation from F_0. Reproduction of R by FF' is shown in condensed form at the right in Table II. Because of redundancies in R and F, this 4×4 residual table contains all of the essential information regarding the reproduction of R by FF'. These residuals make clear that the present asymmetric approach has completely recovered, with but rounding error, the known factor space, and without any recourse to the diagonals of R. The end-result of Tucker's recommended factoring procedure ([9], pp. 118–119), using the same R_{12}, would have failed to do so (cf. [7]).

Although derived in connection with a fictitious example having especially simple features (such as the diagonal form of R_{xy}, TT', and T), the foregoing theory and equations are nevertheless quite general and applicable to any set of empirical data having three identifiable sets of vectors (sub-batteries x and y and battery 2) spanning the same r-dimensional factor space. Thus the order of R must be at least three times as great as its rank. If it were less, then at

TABLE II. ASYMMETRIC SOLUTION AND TEST OF FIT FOR A FICTITIOUS TWO-FACTOR EXAMPLE

	TT'			$F_1=A_1T$			$F_2=R'_{12}F_1(F'_1F_1)^{-1}$			$R-FF'$ (Condensed)			
	5	7		I	II		I	II		1	3	5	7
5	1·556	0·000	1	0·706	−0·707	5	0·802	0·000	1	−0·002			
7	0·000	1·000	3	0·565	0·566	7	0·001	1·001	3	0·001	0·000		
			2	0·706	−0·707	6	0·802	0·000	5	0·000	0·000	−0·003	
			4	0·565	0·566	8	0·001	1·001	7	0·000	−0·001	−0·001	−0·200

	T	
	I	II
5	1·247	0·000
7	0·000	1·000

least some of the diagonals of R would have to appear in R_{xy} or R_{12}, and presumably one of the principal advantages of an asymmetric approach is the avoidance of those diagonals (cf. [9], abstract on pp. 111, 113, 135).

III. FACTORING METHODS

In empirical applications there will of course be problems that did not arise in our fictitious example. The first of these is to determine the rank of R (or of some particular subsection R_{12} that may be of special interest). In effect, this is the question of finding an R_{12} of maximum rank (or of determining the rank of a particular R_{12} if it is of special interest). Some trial-and-error, in which various R_{12}'s are tried, may be unavoidable here, but for a particular empirical R_{12} the problem of determining its rank is reducible to that of fitting it with a factorization of minimum rank.

The simple factoring procedure that was used here, in which a set of r linearly independent columns of R_{12} was used as an A_1, might be useful in preliminary explorations as to the possible rank of a particular R_{12}, and especially as to a lower bound thereof. That procedure could be adapted to extract one dimension at a time, for taking some one column of R_{12} as the first column, $A_{1\mathrm{I}}$, of A_1, the first row, $A'_{2\mathrm{I}}$, of A'_2 could be quickly found by least-squares. That is, given an $A_{1\mathrm{I}}$, the associated post-multiplier $A'_{2\mathrm{I}}$ satisfying

$$R_{12} = A_{1\mathrm{I}} A'_{2\mathrm{I}} \tag{11}$$

with a minimum sum of squared discrepancies is given by

$$A'_{2\mathrm{I}} = (A'_{1\mathrm{I}} A_{1\mathrm{I}})^{-1} A'_{1\mathrm{I}} R_{12}. \tag{12}$$

Here, of course, the matrix $(A'_{1\mathrm{I}} A_{1\mathrm{I}})^{-1}$ is simply the reciprocal of the sum of squared entries in $A_{1\mathrm{I}}$, and the matrix $A'_{1\mathrm{I}} R_{12}$ is just a row vector whose elements are weighted sums of the rows of R_{12}. Those weights are the elements of $A_{1\mathrm{I}}$.

Given an $A_{1\mathrm{I}}$ and its associated matrix $A'_{2\mathrm{I}}$, the residual matrix

$$R_{12\cdot\mathrm{I}} = R_{12} - A_{1\mathrm{I}} A'_{2\mathrm{I}} \tag{13}$$

could be formed, a column of it taken as the second column, $A_{1\mathrm{II}}$, of A_1, and the associated least-squares $A_{2\mathrm{II}}$ computed by

$$A'_{2\mathrm{II}} = (A'_{1\mathrm{II}} A_{1\mathrm{II}})^{-1} A'_{1\mathrm{II}} R_{12\cdot\mathrm{I}}, \tag{14}$$

etc. This process could be continued until the non-vanishing residuals were sufficiently small and randomly arranged to be considered as unimportant errors. At that point the number of columns in A_1 or rows in A'_2 would be the working rank for that R_{12}.

The factoring method just outlined is analogous to the diagonal method of factoring ([8], pp. 101–105) in that it provides perfect fit for those columns of R_{12} and subsequent residual tables from which the columns of A_1 are taken, but at the expense of poorer fit for the remaining columns of R_{12}. This defect could be removed by using row sums of R_{12} and subsequent residual tables, rather than single columns, as columns of A_1. This would make it resemble centroid factoring ([8], pp. 160–170), and, as in centroid factoring, there would be the possibility of reflecting columns of R_{12} and subsequent residual tables. The purpose of such reflections would be to maximize the absolute values of the row sums used as columns of A_1. A slight improvement here might be to use row averages rather than row sums for A_1, in order that the magnitudes in the corresponding A'_2 might be commensurate with those in A_1.

The foregoing 'diagonal' and 'centroid' factoring methods resemble their namesakes only in the formation of A_1. The A'_2 is, in each case, the least-squares 'partner' to the corresponding A_1, for each row thereof, given the corresponding column of A_1, is such as to minimize the sum of squares of the remaining residuals. Thus A'_2 resembles a principal axis factor matrix in its fitting properties ([8], p. 482). It is then natural to ask whether it is possible to have least-squares properties in both A_1 and A'_2. Suppose one started with an arbitrary first column for A_1 (call it A_{110}) and obtained the associated A'_{210} from eqn. (11) by least-squares as

$$A'_{210} = (A'_{110}A_{110})^{-1}A'_{110}R_{12}. \tag{15}$$

Now, using the A'_{210} computed by eqn. (15), one could compute *its* least-squares partner, A_{111}, from eqn. (11) as

$$A_{111} = R_{12}A_{210}(A'_{210}A_{210})^{-1}. \tag{16}$$

But substitution of eqn. (15) three times into eqn. (16) gives

$$\begin{aligned} A_{111} &= R_{12}R'_{12}A_{110}(A'_{110}A_{110})^{-1}\,[(A'_{110}A_{110})^{-1}\,A'_{110}R_{12}R'_{12}A_{110}(A'_{110}A_{110})^{-1}]^{-1} \\ &= R_{12}R'_{12}A_{110}(A'_{110}A_{110})^{-1}\,(A'_{110}A_{110})(A'_{110}R_{12}R'_{12}A_{110})^{-1}\,(A'_{110}A_{110}) \\ &= R_{12}R'_{12}A_{110}(A'_{110}R_{12}R'_{12}A_{110})^{-1}\,(A'_{110}A_{110}). \end{aligned} \tag{17}$$

Now eqn.(17) can be related to something familiar on two counts. First, the expression $R_{12}R'_{12}$ in it is the symmetric matrix

$$H_1 = R_{12}R'_{12} \tag{18}$$

whose principal axis factorization provides the A_1 of Tucker's inter-battery paper ([9], pp. 118–119). Second, the *form* of eqn. (17) is suggestive of the starting point for Hotelling's iterative method (cf. [8], p. 483) for obtaining the principal axis factors of H_1. To see this more clearly, note that eqn. (17) could be written as

$$A_{111} = H_1A_{110}k, \tag{19}$$

where the whole expression

$$k = (A'_{110}R_{12}R'_{12}A_{110})^{-1}\,(A'_{110}A_{110}) \tag{20}$$

is but a constant or scalar multiplier. Equation (19) shows that, except for an unimportant constant multiplier, A_{111} is obtained simply by post-multiplying H_1 by A_{110}. But if this process is continued (i.e., if H_1 is post-multiplied by A_{111} to get an A_{112}, and then by A_{112} to get an A_{113}, etc.), it will converge to a point where two successive A_{11}'s are proportional to each other and to the first principal axis factor loadings of H_1 (cf. [8], p. 483). In the present context, this continuation amounts to finding successive least-squares counterparts, i.e., A'_{2I1} for A_{1I1}, then A_{1I2} for A'_{2I1}, etc., but with no A'_{2I} ever appearing explicitly. Arrival at convergence signifies that no further improvement is possible, whereupon the resulting A_{1I} and A'_{2I} take on the interesting property that each is least-squares counterpart to the other.

The foregoing reasoning can be extended to subsequent columns of A_1 and rows of A'_2, though the details will not be undertaken here. The essence of the extension lies in the fact that if a complete A_1 whose columns are proportional to the principal axis factor loadings of H_1 is substituted for A_{1I0} in eqn. (17), the expression denoted k in eqn. (20) becomes a diagonal matrix. Thus the principal axis factors of H_1 provide least-squares fit, not only for H_1, but also for R_{12}.

Another way of seeing this is to paraphrase a development of Thurstone ([8], p. 482) and write the expression for the general first-factor residual, ρ_{jk}, for the general entry, r_{jk}, of R_{12}, thus:

$$\rho_{jk} = r_{jk} - a_j a_k. \tag{21}$$

Here a_j and a_k are first-factor loadings taken from A_1 and A'_2, respectively, and it should be understood in what follows that all a_j's are from A_{1I} and all a_k's are from A'_{2I}. Squaring the residual in eqn. (21) gives

$$\rho^2_{jk} = r^2_{jk} - 2a_j a_k r_{jk} + a^2_j a^2_k, \tag{22}$$

and summing over all residuals yields

$$\Phi = \sum_j \sum_k \rho^2_{jk} = \sum_j \sum_k r^2_{jk} - 2\sum_j \sum_k a_j a_k r_{jk} + \sum_j \sum_k a^2_j a_k^2. \tag{23}$$

To minimize Φ we set equal to zero its partial derivatives with respect to a particular a_j and a particular a_k, getting

and

$$\frac{\partial \Phi}{\partial a_j} = -2\sum_k a_k r_{jk} + 2a_j \sum_k a^2_k = 0 \tag{24}$$

or

$$\frac{\partial \Phi}{\partial a_k} = -2\sum a_j r_{jk} + 2a_k \sum_j a^2_j = 0, \tag{25}$$

and

$$\sum_k a_k r_{jk} = a_j s_2 \tag{26}$$

$$\sum_j a_j r_{jk} = a_k s_1, \tag{27}$$

where s_2 and s_1 are, respectively, the sums of squared entries in A'_{2I} and A_{1I}. Now when eqns. (26) and (27) are written out for all values of j and k, the results can be stated in the matrix equations

and

$$R_{12} A_{2I} = A_{1I} s_2 \tag{28}$$

$$R'_{12} A_{1I} = A_{2I} s_1. \tag{29}$$

Equation (29) can be solved for A_{21} to be substituted into eqn. (28), with the result, after multiplying through by s_1, that

$$R_{12}R'_{12}A_{11}=A_{11}s_1s_2. \tag{30}$$

But this equation merely states the unique convergence or self-generative property (cf. [8], p. 483, and earlier discussion here) of a column vector A_{11} that is proportional to the first principal axis factor loadings of $R_{12}R'_{12}$, or H_1. Thus again we arrive at the result that a least-squares A_{11} is obtainable from the principal axis factorization of H_1, and the same thing could be shown for subsequent columns of A_1.

IV. OTHER PROBLEMS

Once a particular R_{12} has been made to yield, by some factoring procedure, an A_1 of rank r, there remains the question whether that A_1 can be divided into *two* parts, A_x and A_y, *each* of which is of rank r. The various R_{12}'s should be formed with some foresight for this problem. Here a guide is the requirement that R_{xy} must have rank r. Even so, some amount of trial-and-error on this score might be unavoidable. For this reason, as well as for the purpose of investigating the overall rank of a number of possible R_{12}'s, it might be less costly computationally to factor them initially by a less expensive procedure such as the ' centroid ' method outlined earlier, and to factor only the final R_{12} by the complete least-squares approach. This of course depends upon available computing facilities.

When least-squares factoring is to be applied, some computational savings may be realized by finding A_2 rather than A_1. This is because H_1 must be at least $2r \times 2r$ in size in order that A_1, once obtained, may be divided into the two sections A_x and A_y, each having at least r rows. The least-squares A_2, on the other hand, is obtainable quite analogously from the principal axis factorization of

$$H_2=R'_{12}R_{12} \tag{31}$$

([9], pp. 116–117), which may be as small as $r \times r$ if of full rank. Thus in many empirical applications both the formation and the least-squares factorization of H_2 may involve less numerical work than for H_1. Given A_2, A_1 is readily obtainable by least-squares from eqn. (1) as

$$A_1=R_{12}A_2(A'_2A_2)^{-1}, \tag{32}$$

for here the inverse is diagonal because A_2 is a principal axis factorization (cf. [8], p. 501).

Once the A_x, A_y, and R_{xy} are settled upon, there remains only their use in eqn. (9) to obtain TT'. Formation of A'_xA_x and A'_yA_y and their inversion will test the rank of A_x and A_y, for if either product is not of rank r, its inverse will (as it is popularly put) 'blow up'. Thus large entries in either of the computed inverses constitute a warning as to weak rank r. Conceivably this weak rank could arise merely from scaling down some columns of A_1 to a point where their contribution to the products A'_xA_x and A'_yA_y is overpowered by other columns. This should be guarded against in advance.

A further problem with empirical data arises from the fact that, although the computed TT' is theoretically a symmetric matrix, it will seldom be exactly so in practice. The expression on the right in eqn. (9) is not symmetric in form. Severe departures from symmetry indicate important violations of the theoretical requirement that R_{xy} refer to the same factor space as does R_{12}. Even when R_{xy} and R_{12} have the same rank r, and hence the same *number* of factors, the *nature* of the factors could differ in various ways. A still more insidious possibility is that, though the two sets of factors might differ in nature, they could conceivably interact so as to allow the computed TT' to be fairly symmetric: in other words, symmetry of the computed TT' is a necessary but not a sufficient condition for applicability of the present inter-battery approach. How precisely distortions such as this might arise is hard to see; but usually they should manifest themselves in a poor reproduction of R_{xy} by the appropriate parts of the resulting F.

Minor deviations from perfect symmetry of the computed TT' are probably unavoidable in empirical applications, so that a 'symmetrizing' procedure (to use a word suggested by L. L. Thurstone) is necessary. The simplest possibility would be to add the computed TT' to its transpose and divide all entries in the resulting matrix sum by two. That would amount to treating the *jk* and *kj* elements in the computed TT' as equally fallible estimates of the same quantity, and taking their average.

No special problems should arise in factoring the symmetrized TT' in an empirical study, for r real 'factors' are to be extracted from an $r \times r$ matrix, so that there is no fitting problem. Diagonal factoring should be as good as any in this context, unless some other method would fit in better with subsequent rotational objectives. If the last one or more factors extracted are very weak, and no difficulties were encountered in inverting A'_xA_x or A'_yA_y, it may be a sign that the R_{xy} being used has weak rank r. If it should happen that R_{xy} has rank greater than r, that would create no difficulty in factoring, but it would constitute a violation of the requirement that R_{xy} and R_{12} must refer to the same factor space, so that the resulting fit for R_{xy} would be poor.

Once a transformation T is found, its application as a post-multiplier on A_1 to get F_1 by eqn. (4) is straighforward, whereupon F_2 is obtainable by eqn. (10) or eqn. (5). Equation (5) may be preferable to eqn. (10) for obtaining F_2 if A_2 has already been obtained explicitly, and if T is readily inverted, as it would be if triangular or if a principal axis factorization. If T is triangular, the elements of its inverse (also triangular) can be obtained one at a time using the definition $TT^{-1} \equiv \mathrm{I}$. If it is a principal axis factorization of the symmetrized TT', then it is proportional by columns to an orthogonal transformation ([8], p. 501), and its inverse is simply its transpose with each row divided through by the respective sum of squared elements.

Forming F_1 and F_2 into the complete F is the last step in the present asymmetric factor approach. Tests of fit of any part of R by the appropriate parts of F are now possible, and any type of rotational adjustment that is desired may be applied to F.

V. CONCLUSIONS

This paper has briefly outlined the history of asymmetric solutions in latent structure analysis and factor analysis, and has presented a new asymmetric solution for factor analysis, together with suggestions on how to overcome problems that arise in dealing with empirical data. It is believed that this asymmetric approach will prove useful in practical application, more especially where the investigator desires to avoid the estimation of communalities.

REFERENCES

[1] ANDERSON, T. W. (1954). On estimation of parameters in latent structure analysis. *Psychometrika*, XIX, 1–10.
[2] CATTELL, R. B. (1944). Parallel proportional profiles and other principles for determining the choice of factors by rotation. *Psychometrika*, IX, 267–283.
[3] CATTELL, R. B., and CATTELL, A. K. S. (1955). Factor rotation for proportional profiles: analytical solution and an example. *Brit. J. statist. Psychol.*, VIII, 83–92.
[4] GIBSON, W. A. (1955). An extension of Anderson's solution for the latent structure equations. *Psychometrika*, XX, 69–73.
[5] GIBSON, W. A. (1956). Proportional profiles and latent structure. *Psychometrika*, XXI, 135–144.
[6] GIBSON, W. A. (1960). Remarks on Tucker's inter-battery method of factor analysis. *Psychometrika*, XXV, 19–25.
[7] GIBSON, W. A. (1961). A retraction on inter-battery factor analysis. *Psychometrika*, in the press.
[8] THURSTONE, L. L. (1947). *Multiple-factor analysis*. Chicago: University Press.
[9] TUCKER, L. R. (1958). An inter-battery method of factor analysis. *Psychometrika*, XXIII, 111–136.

Vol. XIV Part 2 The British Journal of Statistical Psychology November 1961

A NOTE ON ROTATION AND PSYCHOLOGICAL SPACE[1]

By R. F. Jarrett
University of California, Berkeley

"*You may clothe a vector space in an inner product; but it's born naked.*"

The purpose of this note is threefold: to consider the conditions under which the concept of distance (both physical and psychological) and angle (the pure number function of distance) are meaningful; to suggest that there is some evidence that these conditions do not obtain throughout psychological space; and to consider briefly the implications of this evidence. My argument will be that distance (and therefore angle) is meaningful only when the space involved possesses the property of dimensional isotropy.

I. Isotropic Space

As I shall use the term, isotropic space is one in which, in principle, measurement operations can be identical no matter what direction one wishes to measure. The long-standing use of standard scores in psychology seems to have misled many psychologists into the uncritical and unrecognized assumption of the isotropy of most psychological spaces[2]. And the question I am raising is whether or not the *dimensions* of various psychological spaces are such that these spaces *can* be isotropic. Consider, for instance, the two-dimensional space defined by the measurements height (cm) and weight (kg) of a sample of adult males. If these pairs of scores be plotted, all psychologists will expect the bivariate scattergram to be elliptical, the major axis being inclined to the horizontal. It is easy to see that, relative to the original reference-variables (height and weight), this major axis has an equation of the form: $a \,.\, x$ cm$+$ $b \cdot y$ kg$=K$ (units unspecifiable). Now suppose someone were to propose that we project each of our points perpendicularly on to this major axis, and, having selected two arbitrary points on that line determining an origin and a unit, that we measure each projected point on this line in terms of that unit. This new dimension will be the one along which we should measure if we wished, by means of a single measurement, to maximize the information concerning individual differences contained in the original set of *pairs* of measurements. Yet no one would consider 'distances' along this 'dimension' meaningful or even practically useful. The reason is that whether or not he has given any consideration to the physicist's *dimension analysis*, it is perfectly obvious that you can not meaningfully add heights to weights. Moreover it is equally clear that were we to convert height and weight to standard scores the corresponding major axis transform is still meaningless, and there is not the slightest possibility of ever being able, independently of measures of height and weight, to determine a ' score ' in this ' dimension '.

---

[1] I am much indebted to Dr. John P. McKee for assistance with the argument and its exposition.

[2] One way of avoiding the difficulties inherent in the use of distance notions is to refuse all recognition to psychological variables. This solution is that of the consistent behaviourist, who is concerned solely with the relations between the input and output events on the two sides of the organism—'a psychology', in Egon Brunswik's terms, 'without the organism'. But the continued vitality of the factor analytic approach, as well as the studies of stimulus generality and of the dimensions of similarity, suggest that not all the problems of interest to psychologists will yield to the former type of approach.

Yet factor analysts seem to believe that for any two psychological tests (I set aside the whole problem of scaling) just such major axis transformations are perfectly legitimate and useful so long as the original scores are first converted to standard scores. This case is as fallacious as the former, and for the same reason.

Consider a hypothetical example. An investigator obtains (i) a measure of religious devoutness, (ii) a measure of the difference limen for monochromatic light of 490 millimicrons, and (iii) a measure of oral fixation. He then plots each of his subjects in the resultant three-dimensional ' space '. In general, so long as he sticks to his original data and so long as he does not measure more than three variables, he is not tempted to rotate his axes because he is immediately struck with the nonsense of a metric that would consist in some combination of the three sets of measuring operations. But two things conspire to mislead the factor analyst. First, he always begins with more than three dimensions, and so can not plot his data. Secondly, he not only converts his data to standard scores, but he then proceeds to correlate his dimensions: that is, he fixes his attention on a matrix formed by the product of the one we have described by its own transpose. In carrying out this arithmetical ritual, he completely loses sight of his original operational dimensions.

In ordinary space, all three dimensions are of the same kind—distance, as measured by yardsticks, calipers, and so on. Though there are many different practical operations for measuring this physical distance, it is possible, in principle, to use only one, and to use that one for each of the three spatial dimensions commonly called length, breadth, and depth. Further, it is possible to use this single operation to measure this same physical distance between points located anywhere in the space: that is, *distance throughout the space has the same meaning even when it is measured in directions which are oblique to any or all of the arbitrary reference axes.* All kitty-corner distances have the same meaning as distances parallel to the axes.

II. THREE SENSES OF THE TERM DIMENSION

To make clear what I mean by *dimensional homogeneity* or isotropy, it is necessary to distinguish between three important senses in which the term ' dimension ' is used. One use is simply to indicate the operations determining one member of a set of coordinates which specifies a referent in its reference space. In this sense, height is an operational dimension; as applied to persons, so are religious devoutness, difference in threshold for a particular wavelength of light, and oral fixation. The chief operational dimensions used in the description of physical objects include length, width, height, weight, volume, temperature, latitude, pressure, and electrical resistance.

One of the assumptions of some brands of factor analysis is that the number of psychological dimensions commonly used to describe the reference objects (i.e. the persons measured) is larger than is necessary for the complete specification of those objects. Hence, applying Occam's razor, we start a search for the minimal number of common reference axes.

This brings us to the second use of the term ' dimension '—to refer to a member of such a *minimal set* of axes. The factor analytic search for a minimal set of psychological dimensions is based on an analogical argument from physical space. Suppose, for instance, that each of a number of points on a line (or on a plane) were to be specified by a set of three coordinates and that the line (or plane) in question did not coincide with one of the reference axes (or with a plane determined by a pair of them). In this case some of the original dimensions used are unnecessary, since an appropriate selection of a smaller set of dimensions would suffice to specify each object uniquely and to describe all the relations among them. For example, each point on the line could be uniquely specified by its distance from an arbitrary origin. Similarly, in the plane, an arbitrary origin and any two reference axes would permit a unique specification of each point in it.

Let me take a further example—specifying the point at the lower southeast corner of the southernmost eave-support of my house. A minimum of three spatial coordinates will be required. The three spatial dimensions are specified each by a distance; and the space defined by these three dimensions is dimensionally isotropic because in principle the two distances from the property lines and the height above the foundation can be measured in the same way.

In this space we can transform the coordinate specification of a point from one system of coordinates into a variety of other systems. So, for example, we learn of rectangular and oblique Cartesian coordinate systems, as well as polar coordinate systems in the two-dimensional space of the plane, and other coordinate systems for the spatial aspects of the ' real world ' could readily be devised if required. Finally, in this isotropic space we may rotate our reference axes to any convenient arbitrary position, because the measurement operations in the various directions are identical. Our familiarity with these facts of everyday space must share responsibility with our use of standard scores for our failure to recognize the fallacies of assuming isotropy where it does not obtain.

In dimensionally non-isotropic space we can use only a limited class of transformations. We may displace the origin, and we may alter the size of the units in terms of which one or more of the reference magnitudes are expressed. But rotational transformations are meaningless because in such a space no meaning whatever attaches to a generalized distance. For differences between points in a non-isotropic space there is no alternative except to use the names of the particular operational dimensions which are involved, whether those dimensions be physical (e.g. duration, mass, length) or psychological (e.g. religious devoutness, oral fixation, difference limen). No generalized metric can be associated with differences in location: we must be content with vectors of differences.

The general class of transformations which we learn in analytic geometry is appropriate in its entirety only to a very special type of space—the dimensionally isotropic space. The notions of generalized distance and angle (the pure-number function of such generalized distance) are appropriate *only* to such a space. Unfortunately ordinary everyday ' space ' is of this special kind; though multi-directional, it is only uni-dimensional.

I turn now to the third sense in which the term ' dimension ' may be used—one that is extremely important. When the layman speaks of time-space as four-dimensional, he means that he needs three spatial coordinates and one temporal coordinate. But there is a real sense in which this time-space is only two-dimensional: viz., the sense that only two fundamentally different kinds of operation are required to determine the four coordinates specifying the system in question. These are the operations of measuring distance and those of measuring time.

It may be objected that this time-space is really three-dimensional inasmuch as any world-point—to use Minkowski's (1913) phrase—may equally well be identified by two angles, one length, and one time: three fundamentally different kinds of operation are involved in measuring distance, angle, and duration. My reply is that it is possible to measure angles in terms of lengths, so that we can analyse the measurement operations into only two fundamentally different kinds. It is this minimal dimensionality which is sought by those varieties of factor analysis which seek unique solutions to the rotation problem.

A second objection is that the two-dimensional time-space may in fact be only one-dimensional: we have not as yet been able to make the proper analysis. Minkowski's demonstration that the Lorentz transformations (introduced to account for difficulties in which physics was involved by the failure to detect the ether-drift expected from Newtonian mechanics) represented simply a rotation of reference-axes in a four-dimensional space seems to suggest precisely such dimensional homogeneity. Minkowski's equations " represented the rotations of the coordinate system in a four-dimensional continuum (space and time combined) with an indefinite metric " (Bergman 1956). Minkowski certainly suggests the dimensional homogeneity of the space-time continuum, though his paper does not make clear just what operations could be used to measure distance along the various space-like (' raumartiger ') or time-like (' zeitartiger ') directions within space-time.

III. THE CRITERION FOR ISOTROPY

The absence of specifiable operations for scaling should not obscure one important point, namely, that *all physical laws which are stated in terms of the space-time manifold of relativity assume the same form, irrespective of the type of Minkowski transformation of the coordinate system that has been used.* It is this which makes the space-time continuum dimensionally homogeneous or isotropic.

Psychology has two alternatives. The first is somehow to demonstrate that such *operationally different* dimensions as religious devoutness, visual difference limen, and oral fixation (or Binet IQ, ascendance-submission, ethnic prejudice and socio-economic status), etc., can nevertheless in principle be measured by the same operation[1]. 'Spaces' defined by such operationally equivalent dimensions would presumably be dimensionally isotropic, and rotation of reference axes within such spaces would correspond to psychologically meaningful operations. The other is to establish laws involving various operational dimensions, and then to determine empirically whether the mathematical forms of those laws remain constant when the reference axes are rotated in whatever tentative 'space' is selected. If they do, the space is isotropic; otherwise, not. In the former case, the notion of psychological distance will be valid. In the latter it will be meaningless, since there are no procedures for demonstrating differences or equalities between such distances.

Attneave (1950), in an interesting paper, has made a preliminary examination of the space defined by judgments of visual differences and similarities, and indeed has anticipated some aspects of my discussion. In considering the hypothesis that psychological space is Euclidean, he remarks that the " most important psychological implication is that, within a space of given dimensionality, one set of axes or one frame of reference is as good as any other. For example, colours are commonly described in terms of brightness and saturation. If psychological space were Euclidean, it would be entirely possible and psychologically proper to rotate these axes and discuss colour in terms of attributes which might be termed 'sightness' and 'braturation', axes cutting across those we generally use " (p. 521). Attneave's own data led him to doubt the Euclidean hypothesis.

So far, psychological factor analysts have tacitly assumed the isotropy of their psychological space. They have thus cavalierly ignored the difficulty which the physicist has in *demonstrating* his isotropy. The physicist has developed his methods for measuring magnitudes, and establishing quantitative relations among his operational dimensions, on the assumption that time-space involved two different fundamental dimensions. Experimental findings revealed the theoretical usefulness of the assumption of homogeneity. But factor analysis, deceived by its standard scores and the arithmetical identity of computing standard scores for its heterogeneous dimensions, has presumed out of hand that psychological space, analogous to the three-dimensional space of the real world, is dimensionally homogeneous. I conclude that factor analysts have been premature in assuming (as they do in some defences of rotation) that the many operational dimensions of psychological space involve only one *fundamental* dimension. I am impressed with the criterion—admittedly extra-logical—that the form of laws remains invariant under all rotational transformations if dimensional homogeneity obtains ; but the psychologist has few quantitative laws to which he can apply this criterion.

---

[1] Cattell's (1946) extensive list of factor analytic studies of personality reveals that in most cases factor analysts have dimly recognized the importance of dimensional homogeneity in selecting tests for analysis. One does not usually find, in any real factor analysis, so heterogeneous a collection of operational dimensions as our illustrative 'religious devoutness', 'difference threshold', and 'oral fixation'.

It seems likely that the search for rotational criteria by factor analysts is related to the problem of dimension analysis, though I have been unable to find any explicit statement to this effect: (see, however, Attneave 1950). Thurstone has emphasized the notion of simple structure as a criterion for rotation. Anyone who would combine such a criterion with the expectation of a unique solution would be committed to the view that psychological space is not isotropic. Here the dimension-analysis approach suggests three possibilities for psychology. The first is that if a particular psychological ' space ' is isotropic there can be no uniquely correct orientation of the reference axes: convenience alone will decide which orientation shall be used. This was the position taken by Thurstone before he began to emphasize simple structure. The second is that the dimensions of a particular psychological space are *not* in principle measurable by the same operations and that there is no single fundamental dimension which pervades it. In this case, a unique and ' psychologically meaningful ' rotation of the reference axes might be sought, but the concept of the communality as a metric-less distance becomes meaningless. The third possible psychological space is one for which there are different sets of operational dimensions; within each there might be dimensional homogeneity, while across sets there might be none. Spearman, with his general and special factors, seems to have been closest to recognizing this possibly. Though such a space might claim the advantages both of partial isotropy and of psychological meaningfulness, I suspect that the notion of a partial isotropy will prove very elusive. In any event, psychology seems so far from demonstrating that IQ, masculinity, ethnocentrism, creativity, psychopathy, etc., can all be measured in the same way that to seek reference axes for ' spaces ' defined by dimensions that are operationally so different is probably fruitless. As a practical matter the factor analysis of such complex ' spaces ' serves the useful end of reducing a mass of data to manageable proportions, and of making some statistical predictions; but it does no more. Meaningful sets of reference axes are not likely to be found in such spaces.

IV. Conclusion

Psychologists have long been concerned with notions of psychological distance. Psychophysicists and modern experimental psychologists as well as differential, clinical, and personality psychologists have found reasons for considering this concept. Although not all have explicitly dealt with distance, a certain invariance notion, closely related to the measurement of distance, has often been implicit in their thinking. Sometimes they find it useful to transform the reference axes of their psychological space—particularly those which entail only displacement or uniform shrinking or stretching of the unit reference vectors. But rotational transformations are legitimate only in spaces that are dimensionally isotropic. By this is meant that the space must be of such a character that, in principle, *identical* operations can be used to measure in each of the infinite number of conceivable directions. Whether or not the conceptual spaces employed by the psychologist possess this property is an empirical question.

Experimental work on stimulus generalization and on the dimensions of similarity suggests that some dimensions of sensory space are *not* dimensionally isotropic. The ultimate criterion is the *formal* invariance, under any transformation of the kind mentioned above, of the *laws* relating psychological variables. At present, psychology has few, if any, magnitudes, and hence few if any laws, whose formal invariance could provide evidence for dimensional isotropy.

The factor-analyst's search for convincing rotational criteria reflects concern about the isotropy of common factor space; but his invariance criteria are not appropriate for testing this isotropy. The possibility of such non-isotropy of common factor space suggests that both those who would rotate arbitrarily and those who would rotate to psychological meaningfulness have truth on their side. The arbitrary rotators are correct in so far as their common factor space contains at least one multidimensional subspace which *is* itself isotropic; the others are correct to the extent that common factor space contains at least two such sub-spaces which are not isotropic with each other. The search for simple structure perhaps reflects concern with the problem of identifying several mutually non-isotropic subspaces. But the ideal of a unique solution for every rotational problem indicates a tacit conviction that the minimal set of reference axes necessary to describe a psychological space will contain no two directions which have the same fundamental dimensionality, and, if they do not, then the notion of a generalized psychological distance becomes meaningless.

REFERENCES

[1] ATTNEAVE, F. (1950). Dimensions of similarity. *Amer. J. Psychol.*, LXIII, 516–556.

[2] BERGMAN, P. G. (1956). Fifty years of relativity. *Science*, CXXIII, 487–494.

[3] CATTELL, R. B. (1946). *Description and measurement of personality*. New York: World Book Company.

[4] EINSTEIN, A., *et al.* (1913). *Das Relativitätsprinzip*. Leipzig: B. G. Teubner.

Vol. XIV
Part 2 The British Journal of Statistical Psychology November
1961

THE MEANS OF SECTIONS FROM A NORMAL DISTRIBUTION

By FRANK SANDON

Formerly Selection Officer, Birmingham Education Committee

The following paper gives a table for the mean abscissae of vertical sections of the normal curve. The need and usefulness of such a table are illustrated by problems from the field of mental measurement.

I. PROBLEM

In both psychological and educational inquiries the investigator has frequently to deal with a group of individuals who have been rated by categories. In many large-scale examinations, for instance, the assistant examiners are allowed to use a scale of literal marks, and the chief examiner may then desire to reduce them to numerical equivalents on the basis of the proportion or percentage of the candidates who are awarded this or that mark. Usually it is assumed—though the assumption is not necessarily true—that the population so marked is normally distributed in respect of the underlying variable, so that the categories form sections of the total distribution (cf. Burt 1943, Sandon 1952). Or again, the investigator may be confronted with two populations, both normally distributed, with the same standard deviation, but with different means (e.g. a group of boys and a group of girls from a given educational area, who have both taken an examination in which one sex is superior to the other, boys, for instance, in arithmetic, or girls in English); and the same line of demarcation has then been applied to both groups, so that different percentages have been selected. The investigator, knowing only the borderline and the two percentages, may wish to estimate the means for the two selected tails.

Formulae for assessing the means of a vertical section or of a tail selected from a normal distribution are already available in the statistical literature (Kelley 1923). But for practical purposes it would plainly be helpful to have a single table of values, instead of having to make an *ad hoc* calculation on each occasion.

II. THE CONSTRUCTION OF THE TABLE OF MEANS

One of the unique properties of the normal curve is the relation between the mean deviation of a section of the total distribution and the difference between the two ordinates, upper and lower, which cut off the section. Thus, if z_1 and z_2 are the ordinates, q_1 and q_2 the areas of the tails cut off by these ordinates, and $\bar{x}$ is the mean deviation of the section

$$\bar{x}=\frac{z_1-z_2}{q_1-q_2}.$$

(see Kelley 1923, pp. 100f.)

The meaning and use of the formula may best be understood by the following example suggested to me by Professor Burt for the distribution of intelligence according to occupational class (see this *Journal*, XIV, p. 11, Table I.) In the table below the figures in the first column give the percentages in each social class. From these the values for q, are obtained by cumulative summation, and the values for z_1, from Kelley's table (*loc. cit.*, Appendix C); the values for q_2 and z_2 are the same as those for q_1 and z_1 in the category just preceding. It is assumed that the entire distribution is strictly normal and that the standard deviation is 15 I.Q. points.

Table I. An Illustrative Example

Class	Percentages	q_1	z_1	$\bar{x}$	Mean I.Q.
I	0·3	0·003	0·00915	3·05	145·7
II	3·1	0·034	0·07545	2·14	132·1
III	12·2	0·156	0·23930	1·34	120·1
IV	25·8	0·414	0·38964	0·58	108·7
V	32·5	0·739	0·32501	−0·20	97·0
VI	26·1	1·000	0·00000	−1·24	81·3

Thus for class V we have

$$\frac{z_1-z_2}{q_1-q_2}=\frac{0{\cdot}32501-0{\cdot}38964}{0{\cdot}739-0{\cdot}414}=-0{\cdot}20\text{ S.D. or }97{\cdot}0\text{ I.Q.}$$

The actual value for the mean, calculated from the detailed test measurements, was 97·8 I.Q. In the case of the other classes, particularly those in which the numbers are very small, the discrepancies are rather larger—as after all was only to be expected with a sample that could hardly be expected to conform precisely to a normal distribution.

To use the table given below (p. 119) the figures for the percentages have first to be converted into terms of standard deviations. Indeed, the table itself is primarily designed for those investigations in which the borderlines are given in terms of actual deviations. The estimates in the table have been calculated from the values given by Pearson and Hartley (1954), Table I; (cf. K. Pearson, 1931, Table II). In these tables X corresponds with Kelley's x, $Z(X)$ with Kelley's z, and $P(X)$ with Kelley's q or p $(=1-q)$. In Table II (p. 119) the means are given to four decimal places. In the case of the most extreme values the figures in the last decimal may be slightly inaccurate.

III. Applications of the Table

To illustrate the type of problem described above, let us suppose that, in the child population with which the investigator is concerned, there is a small

TABLE II. MEAN ABSCISSAE OF SECTIONS OF NORMAL CURVE

Lower Boundary	Width of Section (in standard deviations)								
	0·25	0·5	0·75	1·0	1·25	1·5	1·75	2·0	∞
−3·00	−2·8602	−2·6949	−2·5112	−2·3158	−2·1144	−1·9110	−1·7087	−1·5100	+0·0044
−2·75	−2·6114	−2·4496	−2·2710	−2·0816	−1·8867	−1·6904	−1·4959	−1·3058	+0·0091
−2·50	−2·3627	−2·2045	−2·0310	−1·8481	−1·6605	−1·4723	−1·2867	−1·1065	+0·0176
−2·25	−2·1140	−1·9593	−1·7914	−1·6153	−1·4357	−1·2565	−1·0810	−0·9123	+0·0321
−2·00	−1·8653	−1·7143	−1·5520	−1·3832	−1·2122	−1·0430	−0·8788	−0·7228	+0·0552
−1·75	−1·6166	−1·4693	−1·3128	−1·1517	−0·9900	−0·8316	−0·6798	−0·5377	+0·0899
−1·50	−1·3679	−1·2243	−1·0738	−0·9206	−0·7688	−0·6220	−0·4835	−0·3563	+0·1388
−1·25	−1·1192	−0·9794	−0·8350	−0·6901	−0·5485	−0·4138	−0·2892	−0·1775	+0·2042
−1·00	−0·8705	−0·7345	−0·5964	−0·4599	−0·3288	−0·2066	−0·0963	0·0000	+0·2876
−0·75	−0·6218	−0·4897	−0·3577	−0·2299	−0·1098	0·0000	+0·0963	+0·1775	+0·3894
−0·50	−0·3731	−0·2448	−0·1193	0·0000	+0·1098	+0·2066	+0·2892	+0·3563	+0·5092
−0·25	−0·1244	0·0000	+0·1193	+0·2299	+0·3288	+0·4138	+0·4835	+0·5377	+0·6458
0·00	+0·1244	+0·2448	+0·3577	+0·4599	+0·5485	+0·6220	+0·6798	+0·7228	+0·7979
+0·25	+0·3731	+0·4897	+0·5964	+0·6901	+0·7688	+0·8316	+0·8788	+0·9123	+0·9636
+0·50	+0·6218	+0·7345	+0·8350	+0·9206	+0·9900	+1·0430	+1·0810	+1·1065	+1·1411
+0·75	+0·8705	+0·9794	+1·0738	+1·1517	+1·2122	+1·2565	+1·2867	+1·3058	+1·3288
+1·00	+1·1192	+1·2243	+1·3128	+1·3832	+1·4357	+1·4723	+1·4959	+1·5100	+1·5251
+1·25	+1·3679	+1·4693	+1·5520	+1·6153	+1·6605	+1·6904	+1·7087	+1·7190	+1·7288
+1·50	+1·6166	+1·7143	+1·7914	+1·8481	+1·8867	+1·9110	+1·9249	+1·9323	+1·9387
+1·75	+1·8653	+1·9593	+2·0310	+2·0816	+2·1144	+2·1338	+2·1444	+2·1497	+2·1537
+2·00	+2·1140	+2·2045	+2·2710	+2·3158	+2·3435	+2·3590	+2·3669	+2·3706	+2·3732
+2·25	+2·3627	+2·4496	+2·5112	+2·5508	+2·5740	+2·5863	+2·5922	+2·5948	+2·5964
+2·50	+2·6114	+2·6949	+2·7517	+2·7866	+2·8059	+2·8156	+2·8199	+2·8217	+2·8227
+2·75	+2·8602	+2·9402	+2·9125	+3·0231	+3·0392	+3·0467	+3·0499	+3·0511	+3·0519
+3·00	+3·1088	+3·1854	+3·2336	+3·2605	+3·2737	+3·2795	+3·2818	+3·2827	+3·2831

batch of twins, and that the mean of the batch is only 95 I.Q. points, the standard deviation being 15 I.Q. in both groups (cf. Sandon 1957, 1959). And suppose further that, for the selection at 11+, we pick out the 14 per cent of the total population having the highest I.Q.'s. This implies that the selection is made at about 1 S.D., above the mean, i.e. at an I.Q. of 115. The mean deviation of the children thus selected will be the average of the tail above this borderline. This indicates that this mean will be 1·53 S.D. above the general mean, i.e. the average I.Q. of the selected group will be 123. But for the twins, however, a cut at 115 will be 20 points of I.Q. (i.e., 1·33 S.D.) above their own mean. Interpolation in the table shows that the mean of the selected twins will be at 1·80 S.D. above their own mean, or 1·47 S.D. above the general mean, i.e. at an I.Q. of 122. Hence even the selected group of twins will be, on the average, poorer than the group selected from the population as a whole.

As a further example, let us take an instance from the field of 'subjective probability'. In a recent investigation on the estimation of success in trials depending largely on luck Professor Cohen (1960, p. 140) found that his grammar school pupils estimated their chances of being right in their guesses for certain test problems would be as follows: (relative 'frequencies' given in fractions): Very Unlucky, 0·07; Unlucky, 0·17; Realistic, 0·52; Lucky, 0·17; Very Lucky, 0·07. If we assume the underlying variable to be normally distributed, these proportions would correspond to the following borderlines between successive categories: −1·47, −0·71, +0·71, and +1·47 of S.D. The widths of the corresponding categories will thus be ∞, 0·76, 1·42, 0·76 and ∞. And, on interpolating in the table below we find that the means of the five corresponding sections of a normal curve would be −1·91, −1·04, 0·00, +1·04, and +1·91. We infer therefore that Cohen's subjects on an average, distinguished more clearly between 'Realistic' and 'Lucky' (or 'Unlucky') than between 'Lucky' and 'Very Lucky' (or between 'Unlucky' and 'Very Unlucky'). Moreover, the effect of the adverb 'Very' on the various estimates was far greater when it was applied to the category of 'luck' than it appeared to be with Cliff's subjects when the adjectives thus qualified related to personal qualities; with Cohen's subjects the multiplying factors are 1·91, 1·04, and 1·84, as compared with Cliff's three values of 1·317, 1·254, and 1·323 (Cliff, 1959, Tables XI–XIII).

Editorial Note. As the formula used by Mr. Sandon does not appear in the more familiar textbooks, and as the references cited include no proof, the following proof is appended, based on an early lecture of Karl Pearson.

The equation for the normal curve may be written

$$z=\frac{1}{\sigma\sqrt{(2\pi)}}\exp\{-x^2/(2\sigma^2)\},$$

Differentiating we obtain,

$$dy/dx=-\frac{1}{\sigma^3\sqrt{(2\pi)}}x\exp\{-x^2/(2\sigma^2)\}.$$

Now let $\bar{x}$ be the mean (the abscissa of the centroid) of a vertical slice or section bounded by the ordinates, x_1 and x_2. Then by the definition of such a mean,

$$\begin{aligned}(q_1-q_2)\bar{x} &= \frac{1}{\sigma\sqrt{(2\pi)}}\int_{x_1}^{x_2} \exp\{-x^2/(2\sigma^2)\}x\,dx \\ &= \sigma^2\left[\frac{\exp\{-x^2/(2\sigma^2)\}}{\sigma\sqrt{(2\pi)}}\right]_{x_2}^{x_1} \\ &= \sigma^2(z_1-z_2),\end{aligned}$$

where (as in the text) z_1 and z_2 denote the ordinates at x_1 and x_2, and q_1 and q_2 denote the areas of the curve cut off by those ordinates. If we put $\sigma^2=1$, we then obtain

$$\bar{x}=\frac{z_1-z_2}{q_1-q_2}.$$

In dealing with problems like that described on p. 118 (Table I) and in this *Journal* (XIV, p. 11) research students in psychology at University College have found it simpler to use a conversion-table in which the ' lower boundaries ' and the ' width of the sections ' are given in percentages: the standard measures required are given in the body of the table. As in all such tables, interpolation is needed; and the values so obtained are therefore only approximate. The final and more accurate figures must be calculated directly from the equation just given. Roneo'd copies of the table are obtainable from the laboratory. Mr. Sandon hopes to publish a similar table and other examples in a later issue.

REFERENCES

[1] BURT, CYRIL (1943). Ability and income. *Brit. J. educ. Psychol.*, XIII, 83–98.
[2] CLIFF, NORMAN (1959). Adverbs as multipliers. *Psych. Rev.*, LXVI, 27–44.
[3] COHEN, JOHN (1960). *Chance, Skill and Luck: The Psychology of Guessing and Gambling.* Harmondsworth: Penguin Books.
[4] KELLEY, T. L. (1923). *Statistical Method.* New York: Macmillan.
[5] KELLEY, T. L. (1950). The use of literal grades. *J. educ. Psychol.*, XLI, 490–492.
[6] PEARSON, E. S. and HARTLEY, H. O. (1954). *Biometrika Tables for Statisticians.* Cambridge: University Press.
[7] PEARSON, KARL (1931). *Tables for Statisticians and Biometricians.* Vol. II. Cambridge: University Press.
[8] SANDON, FRANK (1952). The distribution of personality assessments. *Educ. Review.*, V, 53–65.
[9] SANDON, FRANK (1957). The relative numbers and abilities of some ten-year-old twins. *J. roy. statist. Soc.*, CXX, 440–450.
[10] SANDON, FRANK (1959). Twins in the school population. *Brit. J. statist. Psychol.*, XII, 133–138.

Vol. XIV The British Journal of Statistical Psychology November
Part 2 1961

THE GIFTED CHILD

By Cyril Burt
University College, London

The following paper examines certain arguments recently advanced to show that the range of individual differences due to genetic selection is too slight to be of practical importance, and in particular that the abilities of gifted individuals are the result of their early environmental opportunities rather than of their genetic constitution. In reply two new pieces of evidence are brought forward: (i) data demonstrating the wide range of variation to be found among children reared under very similar conditions; (ii) after-histories of persons of high ability whose environment offered an impediment rather than aid to their intellectual development. An estimate is made of the relative frequency of gifted individuals in different socio-economic classes, and suggestions are put forward for improving methods of selecting gifted children of high ability.

I. Current Criticisms

The conclusions I have ventured to draw about gifted children, both in earlier publications and in more recent contributions [2, 3, 4, 5, 7], have of late been subjected to considerable criticism, based, it would appear, quite as much on misunderstanding as on genuine disagreement. Several writers[1] have complained (in the words of one of them) that " the highly complicated algebraic proofs offered by Fisher, Burt, Howard, and others are quite incomprehensible to the average teacher and educationist, and probably to most psychologists. Could not ", it is said, " a few straightforward arguments be given without all these mathematical complications? " Most of the critics follow the line taken by Dr. Fleming, Dr. Floud, Dr. Heim, and Mr. Maddox [11, 12, 14, 16]. The commonest contention is " that the role attributed to heredity by members of the Galton school was greatly exaggerated, and that individual differences in ability or aptitude are due far more to social and environmental influences than to genetic conditions ". This certainly is a criticism that can be met without entering deeply into technical details or embarking on elaborate statistical analyses.

But first of all there is one preliminary objection, raised most frequently by teachers, which applies quite as strongly to the arguments of my critics as it does to my own. " Why," asks Mr. Somerville, " do psychologists treat the problems of the supernormal as if they were of the same general type as the problems of the subnormal ? Both issues seem to be regarded as no more than special aspects of the broader and vaguer problems of 'general intelligence'... We have books on 'the backward child' and books on 'the mentally deficient', but no book on 'the gifted child' as such. It is principally for these reasons that the question has been so widely ignored by the public at large." If Mr. Somerville is thinking primarily of practical problems, I should readily agree[1]. The first task of the psychologist, however, is to deal with the underlying theoretical issues—in particular with the question whether subnormal and supernormal ability are mainly acquired or largely innate; and this demands an inquiry into the assumption of an underlying 'general capacity', commonly termed 'intelligence', and into the degree to which variations in this capacity, whether above the average or below it, appear to be determined by genetic tendencies.

---

[1] See below pp. 171f. The editors regret that it is impossible, for lack of space, to print all the letters received on the subject of the present paper. Of those omitted more than half endorse the main conclusions reached in the articles cited, and are concerned either with minor criticisms or with supplementary inquiries. So far as possible these are also dealt with in the following pages.

Mr. Somerville's view seems to be that geniuses and defectives form rare and isolated groups, standing quite apart from the ordinary population. This is an ancient notion, still widely held by members of the general public. But it is surprising to learn (as both correspondence and examination scripts reveal) that it is a view which is also shared by many teachers, and even by lecturers at training colleges.

The first to point out the manifest errors which such a conception involves was Sir Francis Galton; and it may be helpful to recapitulate some of his simpler arguments here. To begin with, he adduces plausible reasons for believing that, except for occasional pathological cases, the basic differences between individuals are differences of degree. "From the village idiot to such illustrious geniuses, as Shakespeare or Newton, every grade and shade of variation is to be found. And the distribution is wholly continuous, like that for stature and other bodily measurements. There are no gaps, no gulfs." Secondly, he argues, none would deny the occurrence of cases of innate dullness and deficiency; and, since all we know of inherited or inborn characteristics indicates that their distribution is symmetrical, it is reasonable to infer that the high abilities of the geniuses and near-geniuses are also innate.

As stated by Galton, and reformulated by biometricians like Pearson and his coworkers, these elementary facts and the argument based on them raises a strong presumption in favour of the theory that high abilities are mainly innate. But in the absence of more direct empirical evidence, they of course remain very far from conclusive. Let us therefore look a little more closely at the detailed evidence.

II. The Inheritance of High Ability

Galton's principal data consisted of the pedigrees of about 1000 men of genius, which were elaborately analysed and compared. "The general result," he says, "is that one-half of the illustrious men have one or more eminent relatives ". The chance that such a man will have an eminent son proves to be well over 200 times as great as the chance for a parent chosen at random, with similar odds in decreasing proportions for remoter degrees of kinship [13]. To Galton's argument in this general form there is one obvious criticism, repeated by a number of our correspondents, namely, that the opportunities enjoyed by the child of a gifted parent will be far more favourable to the development of high intellectual achievements than those enjoyed by the child of a parent chosen at random. Galton of course anticipated criticism along this line, and attempted to forestall it. However, since his major interest lay in the direct inheritance of ability, he was chiefly occupied with cases in which the children *resemble* their parents; and this naturally invites the retort that environmental influences can produce such resemblances quite as easily as genetic influences.

---

[1] In the reports and memoranda on practical procedures which I prepared as psychological adviser to the London County Council the two sets of problems were in fact treated quite separately. The problems of the 'supernormal' or 'gifted' were dealt with in reports on the scholarship system and on methods of selection for secondary schools.

The Mendelian theory, however, maintains that genetic influences may be responsible not only for resemblances between members of a family, but also for their differences. To the staunch environmentalist one of the most puzzling phenomena is the appearance of extremely able individuals in households where both the intellectual and the material conditions would, one might imagine, condemn every child to hopeless ignorance and inefficiency. Accordingly, instead of looking, as Galton did, for instances of brilliant sons born to brilliant fathers, it will, I suggest, be more instructive to note the numerous instances of brilliant sons born to parents in the humblest circumstances. Carl Gauss, generally acclaimed as " the greatest mathematician of all time ", was the son of a German bricklayer ; Laplace, almost as illustrious, the son of a farm labourer in Normandy; Marlowe, Winckelmann and James Mill were sons of cobblers; James Watt, Whewell, Opie, Lincoln and Carlyle, sons of carpenters; Thomas Cromwell, David Cox, and Faraday, sons of blacksmiths; Defoe and Wolsey, sons of butchers; Luther, Zwingli, and John Knox, sons of peasants; Kant's father was a strap-maker, Tieck's a rope-maker, Franklin's a soap-boiler, Bunyan's a tinker, Kepler's an innkeeper, Cardinal Richelieu's a soldier of the guard; Chatterton's ancestors are described as ' unskilled workers, mostly sextons'. And so one might go on[1].

The detailed accounts of the early development of these and many other budding geniuses fully confirm the inference suggested by this brief list. Unfortunately the illustrative cases commonly described in educational textbooks (cf. [17] and refs.)—Macaulay (I.Q. 180), the son of the governor of a colony who was also a F.R.S., Galton (I.Q. 180), the grandson of a celebrated zoologist and botanist, brought up in a cultured home and taught by an almost equally brilliant sister, John Stuart Mill (I.Q. 190) whose father, leader of the new

---

[1] The data for British persons I have taken partly from Havelock Ellis's *Study of British Genius* [10], but mainly from the *Dictionary of National Biography*. For one or two cases the information is slightly conflicting: e.g. some accounts say Wolsey's father was a grazier, others a wool merchant—presumably he kept sheep and sold both their flesh and their wool—but all seem agreed that he was just " an honest poor man ". For foreign personages I have relied on similar publications from the countries concerned, and especially on Miss Cox's volume in *Genetic Studies of Genius* [9].

Utilitarian philosophy and author of the standard *History of India*, acted as the boy's personal tutor and gave him a profoundly erudite education—these are all cases in which the child was brought up in well-to-do circumstances among highly cultivated relatives and acquaintances, where every opportunity and motive were provided for precocious scholastic progress. Such examples, chosen no doubt because they were the most dramatic that could be found, nevertheless leave the impression that the child's exceptional intellectual attainment may be largely the result of his exceptionally favourable environment.[1] It would, I contend, be far more instructive to examine biographical records where the young genius was brought up in relative poverty by parents who neither enjoyed nor encouraged intellectual interests.

Of the older scientists known to fame Kepler has left fairly detailed descriptions of his early life, which can be checked and amplified by other contemporary documents. His father, who with his wife kept a tavern in a small town in Swabia, was continually running off to the wars: he was, says his son, “ a vicious and quarrelsome vagabond, who treated his family very ill, narrowly escaped hanging, and finally went into exile and disappeared ”. Fortunately, the Dukes of Würtemberg, after embracing the Lutheran faith, had provided a scheme of scholarships “ for children of the poor and faithful ” whose abilities qualified them for a university education; their intention was that the state should produce erudite pastors who could defend the new creed against hostile attacks. In this way young Kepler was enabled to attend advanced courses in theology, classics and mathematics. After graduating at Tübingen at the early age of 20 he was recommended by the senate for a post as teacher of mathematics at Graz, and so was enabled to embark on his investigations and publish the volumes which have secured him his niche in the hall of fame.

Even more instructive is the case of Gauss. He was born in a miserable cottage in Brunswick. His father, a bricklayer, intended his son to follow in his steps and provide for his old age, and consequently “ did all in his power to thwart the boy's intellectual interests, and prevent him going off to acquire an education suited to his talents ”. Fortunately at the local school there was a young assistant with a passion for mathematics. He and the ten-year-old pupil worked together after school hours on algebra and mathematical analysis, questioning the current proofs of the binomial theorem and several of Euclid's demonstrations. When Gauss was only fourteen his remarkable accomplishments came to the notice of the Duke of Brunswick, who paid for his further education. At sixteen he started to develop a non-Euclidean geometry; and at eighteen, so he tells us, he discovered the method of least squares.

---

[1] In support of the view that “ genius is the product of environment ” Mr. Harrison cites the Fellows of the Royal Society—chiefly at its foundation “ members of wealthy and noble families ”, like Lord Brouncker the first president, the Duke of Buckingham and the other seven peers, Sir Robert Boyle, ‘ father of chemistry and brother of the Earl of Cork ’, or else close friends of the Court like Pepys and Evelyn. In point of fact the most noted of the original Fellows were neither noblemen nor ‘ close friends of the court ’; they were men of a humbler origin like Newton and Wren, Oxford professors of puritan sympathies who held chairs from which the royalists had been evicted, and even people like Wallis, the mathematician, who had applied his science to breaking the ciphers of the royalist army. In the next generation, when two earls held the presidency in succession, hardly any work of note was carried out by the Fellows (now said to be nearly all “ gentlemen of noble bloud ”) and the *Philosophical Transactions* were not even published. At this period the ‘ scientific geniuses ’ came from an entirely different social level—like Watt, the son of a carpenter and shipwright, Brindley, the mechanic who had never learnt to spell, and the members of the Lunar Society at Birmingham or the Philosophical Society at Manchester—people such as Priestley, Jenner, Herschel, and the like.

Or, to come nearer home, consider Faraday. His father was a journeyman blacksmith, his mother " industrious but uneducated ". Michael was one of a family of ten. Like them he attended a common day school, but learnt little more than reading, writing, and arithmetic. At thirteen, to augment the family earnings, he obtained a job as errand boy for a bookbinder, who afterwards took him on as apprentice. In this way he had a chance to supplement his meagre education by studying the volumes, particularly those on science, which he found on the shelves of the bookshop. A customer gave him tickets for Sir Humphrey Davey's lectures at the Royal Institution; and here he attracted the attention of Sir Humphrey, who eventually appointed him lecture-assistant.[1]

Even when the future genius is born into a family which can provide both opportunity and encouragement, it not infrequently happens that the path which he follows is quite different from that which might have been expected from his social opportunities or his parents' wishes or plans. James Bernouilli took as his motto *Invito patre sidera verso* (" against my father's will I study the stars "); and the biographies of eminent people are full of similar deviations.

As an example let me choose one who has been described as " the greatest man of science that Ireland ever produced ". William Rowan Hamilton, surprisingly enough, remains unnoticed in Galton's collection, in Havelock Ellis's *Study of British Genius*, and in Miss Cox's *Genetic Studies of Genius*. His early education was supervised by his uncle, a curate at a neighbouring village, who was a pedantic and versatile linguist. The uncle's intention was that William should become a parson and if possible a scholarly divine, and towards this end the first essential was to learn the ancient tongues. At five the boy could translate Latin, Greek, and Hebrew, and recite long passages of Homer in the original, as well as much of *Paradise Lost* and *Regained*. By eight he had added French and Italian, and by ten Sanskrit and Chaldee (i.e. the Aramaic in which much of *Daniel* and parts of *Ezra* are written). At thirteen he claimed that he had mastered one language for every year he had lived; and twelve months later he composed and presented a flowery welcome in Persian to the Persian Ambassador who was visiting Dublin. Just before this, however, he had met the calculating prodigy, Zerah Colburn; and (as he relates) resolved to penetrate the secret of the lad's methods, which Colburn himself did not fully understand.

---

[1] I refrain from citing similar examples from our own contemporaries in the world of science and letters. The stories of politicians who have risen from the humblest origins to cabinet rank—the Thomases and the Morrisons, the Bevins and the Bevans—are too well known to need further mention.

The result, to the uncle's great dismay, was that William now became passionately absorbed in all types of mathematical problems—at first however " purely by way of recreation ". At Trinity College he carried off almost all the available prizes, and obtained top marks and the highest honours in both classics and mathematics. He already detected errors in some of Laplace's algebraic proofs, and had laid the foundation for his discoveries in the theory of optics. When he presented his first memoir to the Royal Irish Academy one of the professors declared: " This young man—I do not say *will* be, but *is*—the first mathematician of his age ". Shortly afterwards at the early age of 22 he was elected professor of astronomy. Nor were mathematics and the sciences his sole pursuits; he still kept up his interest in linguistic subjects, and remained throughout his life a prolific versifier. One could hardly find a more striking illustration of superiority in *general* ability.

Each of the foregoing examples would of itself suffice to refute the view that " the special achievements of men of genius are predominantly the results of the special facilities and stimulation afforded by their environment ". Our universities can show long lists of students whose circumstances must have afforded far better opportunities and stimulation for studying natural science than Faraday's, and who have nevertheless made no comparable contributions to the sciences they studied; and there must have been innumerable schoolboys who have received far more encouragement in their work on mathematics than Hamilton or Gauss, who nevertheless barely got beyond the calculus. At the same time, however, each of the cases I have described demonstrates that, in addition to the boy's own ability and interest, there must at least be some kind of opportunity for acquiring the necessary educational foundations—an opportunity, it would seem, too often the outcome of some happy chance or fortunate condition peculiar to the individual's place and time. How many potential geniuses have failed to develop for lack of such provision it would be rash to estimate.

III. Data from Mental Tests

These biographical particulars will, I hope, supply something of the " plain and concrete evidence " that Mr. Somerville desires. If fuller and more exact confirmation is needed, we must turn to data obtained from more recent research on gifted children tested and studied while still at school. The most obvious way of demonstrating the inadequacy of a predominantly environmentalist hypothesis is to compare children brought up from birth in practically the same material and social circumstances. In the course of our inquiries in various orphanages and residential institutions under the London County Council we encountered from time to time an appreciable number of able boys and girls among the inmates who had been transferred during the earliest years of infancy; and we had free access to the official files containing confidential information regarding their family histories. In spite of the relatively uniform environment the variability of the children taken as a whole was actually greater than that of the general population, viz. 17·3 I.Q. points instead of between 15 and 16[1]. The average I.Q. was only 91, and there was a high proportion of dull and defective youngsters. But about one in fifty had sufficient ability to merit a junior county scholarship although owing to their comparatively poor attainments in English and arithmetic, many of those who deserved a scholarship failed, or would have failed, to pass the customary examination. Of eight children with I.Q.'s over 130, five were known to be the illegitimate offspring of fathers of a comparatively high intellectual and social status; two others were children of fathers in the professional classes who had died and left the mother in a state of extreme poverty. Plainly, in cases such as these, it would be out of the question to explain the high intelligence of the child by the special cultural opportunities furnished by his home environment, since the child's only home has been the institution.

Several critics, however, have objected that " cases taken from institutions would seem to be definitely atypical, and can hardly be regarded as representing the ordinary run of school children ". Accordingly let us turn to pupils brought up in their own homes and taught at ordinary elementary schools. To eliminate so far as possible differences in environmental conditions, we may take a representative sample of children whose parents have all much the same social and economic status. From the surveys of the school population of a typical borough I will select children of fathers belonging to the highest level of what in official reports are so-called 'semi-skilled' workers. Above this level we might occasionally find marked cultural or educative influences emanating from the home; and below it there would be an appreciable number of cases in which poverty, irregular employment, or downright neglect might interfere with the child's early development. The distribution of intelligence among children in this group is shown in Table 1.

TABLE 1. DISTRIBUTION OF INTELLIGENCE AMONG CHILDREN OF SIMILAR HOME CONDITIONS

I.Q.	Number	Percentage
Above 130	1	0·1
120–130	10	1·2
110–120	76	8·9
100–110	279	32·8
90–100	297	34·9
80–90	101	11·8
70–80	73	8·6
60–70	13	1·5
50–60	2	0·2
	852	100·0

---

[1] Among the children attending ordinary Council schools the standard deviation is (as a result of the scaling employed) almost exactly 15. If, however, an allowance is made for children in the same age groups who are either excluded on grounds of mental deficiency or in attendance at private or independent public schools, then the standard deviation rises to rather over 16.

The mean is 96·0, and the standard deviation is reduced from 15·0 to 11·8. Even so, the range is still a good deal wider than we might have expected. If we take the old borderline for junior county scholarships as a criterion, then I should estimate that only 1 or 2 per 1000 in this group would win scholarships as compared with about 2½ per cent in the ordinary population. Nevertheless the group itself is large; and, over the whole county, and during the years in which these and similar surveys were made, quite a number actually succeeded in the scholarship examination or gained equally high marks in the 11 plus examination.

Indeed even in cases where the home circumstances were yet more unfavourable, I have known several pupils who have managed to achieve a highly successful educational career. I have room to cite only a single instance—the son of an illiterate dock labourer and his alcoholic wife, the eldest in a family of five. During the day time, except when at school, he was kept pretty busy running errands for his mother and looking after the family. However, while laid up with a broken leg at the age of seven and a half, he acquired a passion for reading, and later started borrowing books from the public library to read in secret at night. The librarian drew my attention to the boy; I tested him, and found he had an I.Q. of 142. He was transferred to another school, which included what was known as a 'scholarship class'; won first a junior county scholarship, then a University scholarship; and today holds a professorship in science at a Northern university.

IV. INTELLIGENCE AND SOCIAL CLASS

There has of late been considerable discussion about the relative number of gifted children in the different 'social classes' or, as I should prefer to call them, 'occupational groups'. The view which prevails among the majority of teachers is apparently that, "since gifted pupils are commonly the offspring of gifted parents, they are naturally to be expected chiefly in the professional classes or the equivalent ranks in commerce or industry". Social psychologists maintain that this is merely "the superficial effect of the unfair advantages that accrue to members of the privileged and well-to-do classes". Let us therefore inquire into the actual figures.

TABLE II. NUMBERS OF GIFTED CHILDREN IN DIFFERENT OCCUPATIONAL CLASSES

Class	*Occupation*	*Number tested* (1)	(2)	*I.Q. over* 110 (3)	(4)	(5)	*I.Q. over* 120 (6)	(7)	(8)	*I.Q. over* 130 (9)	(10)	(11)
I	Higher Professional	74	3·4	52	70·3	9·6	25	33·8	11·5	15	20·3	27·8
II	Lower Professional	330	15·2	136	41·2	25·1	69	20·9	31·8	23	7·3	42·6
III	Skilled	1127	51·9	260	23·1	47·9	96	8·5	44·2	13	1·1	24·1
IV	Partly skilled	356	16·4	69	19·4	12·6	22	6·2	10·2	2	0·6	3·7
V	Unskilled	285	13·1	26	9·1	4·8	5	1·8	2·3	1	0·4	1·9
	Total	2172	100·0	543		100·0	217		100·0	54		100·0

In the course of school surveys undertaken primarily to locate exceptional types of children—the subnormal at one end of the scale and the supernormal at the other—my colleagues and I have obtained detailed information regarding the occupational status of parents in the case of over two thousand pupils attending London schools—a reasonably random sample. Here it will be convenient to keep to the simple fivefold classification introduced by the Census in 1911. I select it, not because it is necessarily the most satisfactory for purposes of psychological or educational research, but because it is the most familiar to the general reader and the most widely adopted in official or semi-official reports. As the definitions of the several classes are expressed in somewhat general terms, I have drawn the lines of division so that the percentage allocated to each class shall be approximately the same as those reported in the Occupation Tables printed in the last available *Census of England and Wales* (1951).

Class I includes persons in higher professional, administrative, or managerial posts; class II persons in lower professional, administrative, or managerial posts (e.g. teachers in elementary schools and those engaged on commercial or industrial work with much the same earnings); class III includes clerks, typists, shop assistants, and foremen as well as skilled manual workers in various trades; class IV consists of ' partly-skilled ' workers; and class V of the remainder—chiefly various types of unskilled labour. In Table II I have given figures for three grades of supernormal children—viz., those forming the top 25 per cent, the top 10 per cent, and the top 2½ per cent of the school population. These correspond fairly closely with those having I.Q.'s of 110 ,120, and 130, or upwards. I choose these proportions, because 2½ per cent represents roughly the proportion of junior county scholarship winners in the early days of such examinations ; they form in fact the type of 'gifted children' for whom we have the most complete records, including a large number of after-histories reaching almost to the middle of adult life; 10 per cent was approximately the proportion allocated to so-called ' central schools '; and 25 per cent represents the average proportion who nowadays pass the 11 plus examination for entry to a grammar school.

Looking along the first row, it will be seen that, out of the 2 172 children who comprised the entire sample, only 74 (barely 3½ per cent—col. 2) were drawn from the higher professional class. Of these all but half a dozen had I.Q.'s above 100; 52 (over 70 per cent—col. 4) had I.Q.'s above 110, and were therefore of grammar school type. Looking down the columns, we see that out of the 543 children of this grammar school type, only 52 (less than 10 per cent—col. 5) were drawn from the higher professional class. Among these 52 there were 25 with I.Q.'s above 120, and 15 with I.Q.'s above 130: thus one in five (20 per cent) of the children in this small subgroup possessed this exceptionally high level of ability (col. 10). Nevertheless, even at this high level it is not Class I but Class II (col. 11) which furnishes the biggest quota; and owing to their great numerical superiority, the three ' working classes ' (III, IV and V) actually supply a larger proportion of children with I.Q.'s over 130 than the higher professional class.

I fancy much of the controversy arises because the various disputants fail to distinguish between the two types of percentage—namely, those reached by comparing the numbers by rows (i.e. occupational classes) and those reached by comparing them by columns (i.e., I.Q.'s). Defining, for the sake of argument, the term ' gifted ' to mean an I.Q. of 130 or more, one could quite truthfully say that a child from class I has a one-in-five chance of being ' gifted ', whereas a child in class II has only one chance in fourteen, and a child from class IV barely one in two hundred. But it is equally true to say that only about a quarter of the ' gifted ' children come from class I. Unless the two statements are carefully worded and explained, they sound contradictory, and may easily lead their respective spokesmen to imagine that their opponents are misinterpreting the facts.

VI. THE ASCERTAINMENT OF THE GIFTED

Assuming then that these figures may be accepted as plausible estimates and that the differences which they reveal are due largely to genetic factors, some kind of selection becomes essential if (in the words of the Act) we are to provide each child with " such instruction and training as may be desirable in view of his age, ability, and aptitude ". Up to now, as our various after-histories show, at least a third of the pupils in the higher intellectual categories have failed to obtain the type of education which their ability seemed to deserve. Not all such failures are to be ascribed to faulty methods of selection. Motivation is as important as intelligence; and development after 11 plus as important as development before 11 plus—and far harder to predict.

Naturally those who believe that genetic differences are too slight to produce appreciable differences in actual development, and therefore attribute the individual's educational and social achievements chiefly to the advantages conferred by his early environment, would prefer to abolish all schemes of selection. In this country, however, the majority of teachers and educationists have been forced by their own experience to recognize the wide range of innate differences, and consequently appreciate the need for selection at both ends of the scale—for the gifted as well as the defective, for the supernormal as well as the subnormal. Nevertheless even they have freely criticized what they take to be the grave shortcomings of existing procedures. The commonest target of attack is the ' intelligence test ' and particularly its use at 11 plus.

The most uncompromising criticisms, however, emanate from the social psychologists. Dr. Floud and her collaborators maintain that the grammar school tends to become " an undisputed preserve of the middle class ". As they rightly point out, the chance that the son of a non-manual worker (e.g. a clerk) will enter ' a secondary school of the grammar type ' is three times that of a working-class boy. And such differences they attribute largely to the reliance on the ' intelligence tests ' introduced by psychologists for this purpose and to the influence of environment on the test-scores [12, p. 143]. Dr. Stott's oft quoted protests [18, p. 27 f.] are stronger still: " educational administrators," he tells us, " clamour for I.Q.'s; . . . the legend that the intelligence test measures innate unalterable intelligence[1] was fathered by an earlier generation of psychologists and mothered by administrators only too glad for someone else to carry the baby "; and, since both the educational administrators who " clamour for I.Q's " and the educational psychologists who construct the tests are themselves commonly drawn from the privileged classes, the results obtained tend to select those who, by virtue of their home background and their individual interests, are likely to turn into typical members of the same social class. " The notion of selecting gifted children by administering written tests of intelligence devised for the 11 plus examination is," we are assured, " hopelessly out of date; not only the statistical evidence, but still more the observational data show that their effect is merely to favour a specially privileged type of child—in short, a child in most cases of the educated middle-class type of parent." And a comprehensive survey of the more popular arguments, leading to the same conclusion, has been set forth by Mr. Flann Campbell [8].

Now in point of fact the introduction of intelligence tests for selecting entrants to grammar schools had a very different history from that which these writers assume. The tests to which they refer were in the first instance devised not for practical purposes but for scientific surveys. To gather data on an extensive scale written tests applied in groups were necessary; and accordingly Mr. R. C. Moore and I [1] developed a number of verbal and non-verbal group tests—opposites, analogies, reasoning, code tests, matrix tests, completion tests—types of test which still form the bulk of the so-called 'intelligence tests' used in the 11 plus examination. The main purpose was to eliminate, so far as possible, the effects of environmental differences, and to provide some kind of comparable scale by which the estimates furnished by teachers from different schools with different subjective standards could be objectively equated. In any case it was always intended that the test marks themselves should be submitted to teachers who knew the individual pupils, so that doubtful or discrepant cases could be re-assessed by individual tests and other means.

In Britain tests of this kind were first employed for educational purposes in a series of experimental studies which I carried out as the Psychologist to the London County Council from 1913 onwards. At that time the Council awarded a very limited number of scholarships annually to pupils of 11 plus who showed exceptionally high ability, and in doing so was guided by the pupils' performances in a purely scholastic examination consisting of two papers only—English and arithmetic. An analysis of the results showed that the number of 'gifted children' thus selected varied enormously from one electoral division to another. In Lewisham, Dulwich, St. Pancras North, and Hackney North, the proportion of scholarships gained in 1913 per 1,000 children in attendance at the elementary schools was well over 7 as compared with an average of 2·9 for the whole county; in Bethnal Green, Haggerston, Lambeth North and Southwark West on the other hand, it was less than 1·0, and in Finsbury only 0·2

---

[1] This statement is based on a widespread misunderstanding. The view adopted by the 'early generation of psychologists' who introduced such tests was that there is an innate general factor influencing mental efficiency in all directions, and the word 'intelligence' was borrowed as a convenient shorthand label. Tests were constructed to aid in assessing this factor. 'Intelligence' has now become a term of popular parlance with a far broader meaning, and, of course, in that sense is by no means unalterable. The imperfections of tests were emphasized from the very outset, and the fact that 'intelligence' *as measured by these fallible methods* alters appreciably was demonstrated again and again (see [2], pp. xv, 151 f, 175).

In the inquiries that followed, the application of intelligence tests such as I had already compiled for researches carried out at Oxford and Liverpool indicated that able pupils coming from the poverty-stricken and semi-illiterate homes in the areas just mentioned were badly handicapped in the scholastic type of examination, particularly in the English paper. Later I was invited to carry out similar investigations with intelligence tests in Bradford, Birmingham, and Northumberland (in the last case with the welcome collaboration of Godfrey Thomson). In London we concluded that the wisest plan would be, not to make a separate paper consisting solely of intelligence tests, but to incorporate a number of problems of that type in the existing papers. This procedure, we believed, would help to frustrate the inevitable efforts at special cramming; whereas the addition of a separate paper called an 'intelligence test' would, we felt sure, give rise to many misconceptions—as has indeed proved to be the case with the 11 plus examinations.

So far from the new test-questions favouring the child of the 'educated middle-class type of parent', as our critics allege, the demonstrable effect was to reduce the disproportionate number of scholarships going to the 'privileged few'. Thus in the sample borough in which our most intensive surveys were carried out the proportion of scholarship winners coming from so-called working class homes (those of skilled, semi-skilled and unskilled manual workers) rose from 1·8 per cent to 3·5 per cent.

But, contrary to the prevailing notion, it is not the proper function of a test applied at 11 plus to pick out children of the highest *innate* ability. If at this age a pupil is to be accepted as a fit candidate for a grammar school, he must also possess a certain modicum of elementary educational attainments. The detection of innate ability is a task that should be undertaken at a much earlier age. With youngsters just up from the infants department an intelligence test, rightly selected and rightly applied, can provide a valuable means of discrimination. What is wanted at 11 plus is no longer a test to discover differences that are inborn, but rather an assessment of the child's present all-round competence. Some writers of course use the word 'intelligence' in that broader sense. But much misunderstanding would be avoided if this equivocal term were dropped in favour of some non-committal designation, e.g. 'general classification test'. As it is, many of the brightest children, particularly those who live in poor areas and attend schools where most of the other pupils are dull and backward, are apt to end up at 11 plus with attainments in English and arithmetic well below grammar school requirements.

One of the earliest results of mental testing was in fact to demonstrate how grossly the abilities of the more gifted pupils are underestimated. Of children with I.Q.'s over 130 as many as 54 per cent—well over half—had educational quotients under 120. Of those with I.Q.'s over 120 only 11 per cent had E.Q.'s over 120, and 6 per cent had E.Q.'s below the general mean. On an average " the abler children are in effect deprived of more than half the educational equivalent of their natural advancement—over 10 per cent of their mental age " [2, pp. 177f].

Nevertheless I fully agree that the test-papers in actual use could be greatly improved. Dr. Stott rightly complains that different psychologists and different tests often produce widely varying I.Q.'s for one and the same child. Part of the trouble arises from the fact that all too frequently the tests employed are never adequately validated. It looks so easy to construct an ' intelligence test ' for the 11 plus examination by inventing sets of opposites, analogies, and problems involving completion, codes, or syllogistic reasoning; and the local authority's examiner no more thinks of trying them out beforehand than he would think of trying out his questions in English and arithmetic.[1] And even tests that have been adequately constructed and validated may have very different standard deviations. Thus a child with an I.Q. of 125 on Dr. A's scale may have an I.Q. of 135 on Dr. B's. An education official familiar only with Dr. A's would infer from B's mark that the child was the brightest in 200, while an official familiar only with B's scale would infer from A's mark that the child was merely the brightest out of 10 or 11. Those who discuss I.Q.'s should always state what particular test-scale they have in mind, and not refer, as they so commonly do, to ' *the* I.Q.' as though all acceptable tests would give the same figure.

But there is a further defect. The tests which are easiest to construct and standardize are tests which turn on logical analysis, aided by a retentive memory —particularly those which can be formulated in verbal terms. Now a large proportion of highly gifted persons carry out their thinking by non-verbal means, and belong to what I have termed a synthetic, intuitive, or productive type rather than to the analytic, logical, or reproductive. They are quick at seeing problems in the concrete, and they reach their conclusions by implicit thinking rather than explicit [1]. Among adults this is true, not only of artists, musicians, inventors, and technologists, but also of nearly all who are marked by unusual originality and creative capacity. As I have pointed out elsewhere [6, pp. 19f.] an excessive analytic bias was a serious fault in Terman's method of selection. The gifted individuals picked out by his tests were in later life conspicuous chiefly for their achievements in the world of academic science: there was not a single painter, sculptor, or musical composer with a reputation comparable with those of the dozen most eminent scientists. Even in their literary productions their success lay rather in academic textbooks and scientific or technical articles than in poetry, *belles lettres*, or other branches of imaginative writing.

---

[1] In compiling a group test for the examination in one educational area the examiner decided to use an ' opposites test ' of his own: to avoid previous familiarity he excluded all the items that appeared in his model (test 21 in *Mental and Scholastic Tests*). The result was that most of the items he included consisted of those which we had thrown out in the course of our preliminary item-analysis because of their low validity.

I believe then that if improved methods of testing and checking tests were introduced, if special abilities and disabilities could be taken into account, and if character and motivation could also be assessed (on the basis for example, of teachers' reports), the accuracy of the selection—the 'predictive value of the examination' as Mr. Harrison terms it—would undoubtedly be increased. However, the shortcomings of the present methods of selection have, I fancy, been much exaggerated. In the inquiries carried out by the National Foundation it was estimated that every year about 78 000 are 'wrongly' allocated; and the figure has been pounced on by the critics. But, as the investigators point out, comparatively few of the 'wrong' allocations are *seriously* wrong. In any case the inevitable fluctuations in growth and interest and the unforeseen accidents of the future put *perfect* accuracy quite out of the question. The real object of the system of selection is, or should be, to make certain of catching in the net the ablest pupils of all—say the top 3 per cent. If at eleven plus you select only 3 per cent, you will miss more than half of them. To be reasonably sure of including all of them when your validity coefficient is only 0·80 or thereabouts, you must select the brightest 25 per cent according to your tests. You will then probably miss only 1 in 400 or 500 out of the brightest 3 per cent.

So far, the types of gifted children I have been discussing, namely, pupils with I.Q.'s of 110, 120, or 130 and over—are by no means rare, though somewhat infrequent; they comprise roughly speaking, those who are the ablest out of 4, out of 10, or out of 50 respectively. Galton's definition of a genius, however, was the ablest person out of 4000, which means one with an I.Q. of 160 or more. Such an individual would differ as much from a first class honours candidate as the latter does from the average skilled workman. And here, I think, the point to underline is the increased spacing in the differences between individuals as the individuals depart more and more from the mean. Suppose we mark a random sample of a hundred children on a uniform scale from 0 to 100. The difference between the 50th and the 51st would be almost imperceptible (50·0 marks and 50·5); but the difference between the brightest of all and the next brightest would be 10 times as great, namely, 96 and 91 marks. It is this increasing rarity towards the top which makes the greatest geniuses—the Shakespeares, the Raphaels, the Newtons, and their like—stand out as almost inexplicable miracles of nature.

Now the interest of the educationists has been concentrated for the most part on selecting and educating pupils with relatively moderate degrees of giftedness. But our richest source of intellectual power lies in those who are the most exceptionally gifted; and these really need an education quite distinct from that provided by the ordinary grammar school or college course. In the past such an education has, in rare instances, been provided by a parent (as in the case of John Stuart Mill); more frequently it has been worked out by the gifted individual himself (as in the case of Galton), aided in certain cases by one or two early strokes of luck (as in the case of Faraday). In a country like the United Kingdom, with a population of about 50 million, there must be at least 12 000 geniuses, in the sense defined by Galton. I estimate that the talents of more than half of them are virtually wasted—largely owing to the inadequacy of our methods of selection and training.

Mr. Harrison complains that " the educational system Professor Burt would have us establish implies a society in which class-distinctions, instead of being reduced or superseded, are sharply accentuated: according to his view it would mean a community arranged in a hierarchy of *intellectual* grades, presumably with a glorified public school for the intellectual élite at the top; according to the view accepted by most contemporary sociologists it would create or perpetuate a community arranged in *social* grades based on income, privilege, or sheer intellectual snobbery ".

There is confusion as well as misconception here. In suggesting " something like a new type of public school " for those endowed with an I.Q. of over 160 my aim was to avoid the constant sense of frustration and maladjustment that so many of them experience while living at home or attending a school of the ordinary type whether primary, secondary, grammar, or public.[1] They need the companionship of those who are their intellectual equals among children and the discipline of those who are their intellectual equals among adults. But it was certainly not my wish to transform public schools into institutions for an " intellectual élite ". Naturally, in reporting results of researches on tests of intellectual ability, I have made little or no reference to the non-intellectual aspects of the problem; and among educationists generally I fancy a preoccupation with selection by examination has led to a similar narrowing of their scholastic outlook. Critics of the public schools and of " class distinctions in the educational system " seem to believe that " professional parents who prefer to send their children to public schools attended by boys from the same exclusive social class as themselves " do so in order to " purchase a superior type of instruction given by a superior type of schoolmaster with only a dozen boys in a form instead of thirty or forty". This is a common, but to my mind a very inaccurate, picture. The reason why the parent from the professional classes sends his son to a public school, often at the cost of drastic financial sacrifice, is not so much a desire to secure a better type of intellectual education (in many respects the intellectual education is often inferior to that given in the nearby secondary schools established and equipped by the local authority) it is rather that he wants the training in character, manners, speech, and social ideals which, with few exceptions, only the better type of public school at present seems able to impart.[1]

---

[1] As is evidenced by the autobiographies of so many of them: cf. [19], for example.

To my mind the disparity in traditional ideals is by far the most important difference associated with class distinctions. As those we questioned continually pointed out, the professions have always prided themselves on their high ethical and social ideals handed down from the past. The medical profession has its code of unselfish service which allegedly goes back to Hippocrates and is enforced by its Council; the legal profession claims to stand for strict and impartial justice; the clergy naturally aim at upholding moral standards; the teaching profession seeks to include among its aims—and has done since the days of Plato—both character training and intellectual integrity, inculcated by example as well as by precept: and finally in this country both politicians and civil servants inherit a tradition of loyalty and public spirit. The code of the average professional family claims to incorporate these various ideals and naturally such parents want their sons and daughters to associate with children from families who cherish the same high principles. In my view these parental aspirations should also be taken into account by our methods of selection.

Be that as it may, we still know too little about the needs and potentialities of gifted individuals at successive stages of their mental growth and of the various ways in which their abilities, interests, and temperamental qualities develop or mature. What is urgently needed is a fresh series of longitudinal case-studies, starting with an intensive investigation of a broad sample of gifted children while still at school, and resumed at intervals until the children have attained the middle of adult life. The inquiries carried out in the London County Council's schools may serve to indicate what is practicable.[2] But both the numbers studied and the information sought were far too limited, and referred to conditions very different from those obtaining today. If the investigation is to last for some thirty or forty years, it should be undertaken not by some single individual, however enthusiastic, but by some scientific or official body. At least one of the investigators should be a member of the local authority's staff: only in this way will they be able to secure the requisite cooperation from teachers, school doctors, social workers, and parents, and to obtain the necessary facilities for carrying out physical, psychological, and educational tests and for studying the confidential data already obtained. The knowledge achieved by such an inquiry would amply repay its cost.

---

[1] This type of motive emerged very clearly in an investigation carried out by Miss J. Whittaker and me into the reasons given by parents in seeking particular types of school for their children: it appeared most conspicuous among parents of the professional class.

Mr. Harrison also objects that I " appear to claim for the gifted group a complete freedom from every type of blemish, physical as well as mental ". This presumably is meant as a rhetorical exaggeration. It should, however, be emphasized that I was making no ' claim ', but merely reporting ascertained facts—viz. that the percentages showing physical defects are smaller among the gifted group than among the general population. But I mentioned one striking case in which the percentage was actually higher—namely, visual defect (chiefly, at the age in question, moderate myopia). Terman's figures show the same difference (20 per cent more or less defective among the gifted, only 16 per cent among the control group; 11 per cent of the gifted wearing glasses, only 5 per cent among the control). Terman does not, however, realize the problem presented by this odd exception to the general rule. How is it to be explained? Has short sight acted as a reason for preferring books and school work to games, or is the difference the result of natural and sexual selection?

[2] Many local education authorities have now been using standardized 'intelligence tests' in the 11 plus examinations for fully ten of fifteen years. It should therefore be possible to make comparative studies (similar to those carried out on a more limited scale in London) of pupils who have passed or failed. The after-careers of large numbers in various parts of the country, both urban and rural, could be followed up at least to the beginning of the University stage, i.e., to the age of 18 or thereabouts. May we hope that psychologists or research-students will be encouraged to collate and analyse the vast mass of data thus available, and, where possible, obtain the comments of the teachers who still remember the pupils in question?

REFERENCES

[1] BURT, C. (1911). Experimental tests of higher mental processes. *J. exp. Pedag.*, I, 93–112.

[2] BURT, C. (1921). *Mental and Scholastic Tests.* London: P. S. King.

[3] BURT, C. (1943). Ability and income. *Brit. J. educ. Psychol.*, XIII, 83–98.

[4] BURT, C. (1959). The examination at eleven plus. *Brit. J. educ. Stud.*, VII, 99–117.

[5] BURT, C. (1959). Class differences in general intelligence. *Brit. J. Statist. Psychol.*, XII, 15–34.

[6] BURT, C. (1961). The gifted child. *Eugenics Review*, LIII, 17–20.

[7] BURT, C. (1961). Intelligence and social mobility. *Brit. J. statist. Psychol.*, XIV, 3–24.

[8] CAMPBELL, F. (1956). *Eleven Plus and All That.* London: Watts.

[9] COX, C. (1926). *The Early Mental Trails of Three Hundred Geniuses.* Stamford: University Press.

[10] ELLIS, H. (1904). *A Study of British Genius.* London: Hurst and Blackett.

[11] FLEMING, C. M. (1948). *Adolescence.* London: Routledge and Kegan Paul.

[12] FLOUD, J. E., HALSEY, A. H., and MARTIN, F. M. (1956). *Social Class and Educational Opportunity.* London: Heinemann.

[13] GALTON, F. (1869). *Hereditary Genius.* London: Macmillan.

[14] HEIM, A. W. (1954). *The Appraisal of Intelligence.* London: Methuen.

[15] HOWARD, M. and BURT, C. (1957). The relative influence of heredity and environment on intelligence. *Brit. J. statist. Psychol.*, X, 99–104.

[16] MADDOX, H. (1957). Nature-nurture balance-sheets. *Brit. J. Educ. Psychol.*, XXVII, 166–175.

[17] PINTNER, R. (1924). *Intelligence Testing.* London: University of London Press.

[18] STOTT, D. H. (1956). *Unsettled Children and their Families.* London: University of London Press.

[19] WIENER, Norbert (1953). *Ex-Prodigy: My Childhood and Youth.* New York: Harper.

Vol. XIV The British Journal of Statistical Psychology November
Part 2 1961

NOTES AND CORRESPONDENCE

THE STRUCTURE OF THE MIND. I

A Note on Physical and Psychological Space

By J. R. SMYTHIES

Department of Psychological Medicine, University of Edinburgh

Professor Burt has brought a number of objections against my formulations of the representative theory of perception and of the nature of the mind-brain relationship. He says first of all that, in common with Dr. Levy, I have " taken the highly ambiguous concept of ' space ', and have applied it to very disparate sets of phenomena." The chief ground for this criticism is apparently that Dr. Levy has suggested that his own ' all-embracing space-time ' may be regarded as a kind of ' factorial space '. This leaves him open to the objections that Professor Burt raises. However, I have never myself suggested that either perceptual space or the space that includes both physical space and perceptual space, are factorial spaces. In my book[1] I based the concept of space on the relations between after-images and other sense-data, that is, on something that is given in experience and depends in no part on mathematics. Indeed Nicod[2] bases geometry itself on his observation of such relations as these.

Professor Burt then goes on to question my use of the concepts of ' sensible ' space and ' physical ' or ' scientific ' space. Each individual, he says, has not just one ' sensible ' space, but several—e.g., a visual, an auditory and a tactual space, each separate; and moreover these are not so much ' spaces ' as ' fields ', i.e., limited in extent. Secondly, he states that my concept of ' scientific ' space is really a ' psychological construction ' rather than equivalent to the ' sophisticated concept of modern physics '.

These difficulties may best be avoided if we pay more attention to the entities concerned instead of basing our discussion on the difficult concept of ' space '. If we do this, we can make the following analysis of the underlying situation. Each individual consciousness manifestly consists, as Professor Burt notes, of a collection of events ; and some of these events are organized into fields—e.g., the visual field of visual sense-data and the somatic field of somatic sense-data. These are *spatial* fields : the individual units which can be isolated from the total field concerned manifestly bear spatial relations to other units in the same field. However, this is not the whole story, as Professor Burt observes here and elsewhere[3]; for there are other contents of consciousness, such as olfactory and gustatory sense-data, as well as thoughts and feelings which do not have spatial properties. Nevertheless it is still true to say that *some* mental entities have spatial properties (or that some mental events have spatio-temporal properties). And it is this key fact about the nature of the mind that the Cartesian theory fails to recognize.

The next step in my defence is to make it clear that my theory does not really deal with the space of physics. It deals rather with the physical objects of common-sense—chairs, houses, trees, etc., and with their mutual spatial relations. The theory can be put simply as follows:

---

[1] J. R. Smythies, *Analysis of Perception*, 1956.

[2] Jean Nicod, *The Foundations of Geometry and Induction*, 1930.

[3] Cyril Burt, *J. Soc. psych. Research*, XLI, 1961, pp. 55–60.

1. Physical common-sense objects bear spatial relations to each other, and they are located in one common spatial continuum; that is, any one object could, in principle, *touch* any other physical object.

2. One man's visual sense-data bear spatial relations to each other and are located in another continuum—his visual field in his consciousness ; that is, each of his visual sense-data is contiguous with other sense-data or with other visual sense-data which in turn are contiguous with other visual sense-data, and so on; and this process exhaustively accounts for all the sense-data in his visual field. It will be noted that we here depend on basically simple topological relations, for it is these that the contents of consciousness manifest most clearly.

3. The essential step now is to ask what are the relations between sense-data and physical objects. Clearly they are not spatially contiguous with each other in the way that physical objects touch other physical objects, and that sense-data are contiguous with other sense-data. Professor Price's theory answers this question by saying that there are no spatial relations between sense-data and physical objects, but only causal relations ; the total universe of events consists of a physical world of common-sense objects, and, in addition, a number of mental worlds—one for each individual. It is in our description and explanation of the *behaviour* of common-sense objects that the concept of the space of physics arises. Each mental world consists of that individual's sense-data (both the spatial and the non-spatial kind), images, thoughts, etc. Thus consciousness is not *merely* a ' special type of relation or interaction '[1]. It does not consist *solely* of my awareness of sense-data, although, of course, this awareness is an essential feature of consciousness. Included in the class of events that comprise each individual consciousness are the sense-data themselves and their behaviour. Professor Broad's theory differs from Professor Price's theory only in postulating that the physical world and the various mental worlds form different cross-sections of a common n-dimensional manifold.

In my own analysis the problem of the relation between the visual field and the other components of consciousness becomes no longer a matter of mind–brain relations, nor one for physiologists to worry about ; it becomes a problem for pure psychology with introspection as our only source of information.

I will conclude by dealing with two minor points. (i) Physicists have only assumed that events " which act as causes and the results which ensue as their effects are and must be events in the same space ", because they have never supposed that there might be other space-time systems *outside* the confines of the world of physical objects. But, as Ducasse has shown,[2] there is nothing in the logic of causality that permits us to state on *a priori* grounds that events in one space cannot be causally linked with events in another space or even with non-spatial events. Causality in this context means no more that whenever events a, a^1, a^2, . . . a^n occur in the brain, then events b, b^1, b^2, . . . b^n, occur in the owner of that brain's consciousness. (ii) My use of television as a model for the representative theory of perception specifically excludes any need for the detail of the mechanical arrangements (such as scanning) to be the same in both cases. The model is a logical one, and illustrates the essential features of a non-symbolical representative mechanism such as I suppose the human visual mechanisms to be.

---

[1] Cyril Burt, *loc. cit.*

[2] C. J. Ducasse, *Nature, Mind and Death*, 1951.

THE STRUCTURE OF THE MIND. II

To the Editor, British Journal of Statistical Psychology

Sir,—I should like, if I may, to add my support to the case made out by Dr. Taylor, Mr. Hansel, and other contributors to your pages for a purely objective psychology. Such a psychology, as they rightly argue, should be based on the same essential concepts and adopt the same empirical methods as every other branch of natural science. It would, as Dr. Taylor more especially has maintained, keep well within the bounds of what he calls 'Newtonian science', it should retain a rigorously mechanistic or deterministic attitude, since that is the only way to avoid the illegitimate assumption of subjective qualities and occult entities, such as 'ego', 'psyche', 'consciousness', 'mind', and the other elusive concepts so dear to the philosophical psychologist.

Professor Burt, if one may judge from his recent papers, would have us resurrect the traditional Cartesian notion of a simple unextended substance, called a mind or a soul. He seems to take it for granted that the concept of mind is an empirically established theory. In that case one is bound to ask what exactly is the observational evidence on which the proof of the theory is presumed to rest. In the view of most psychologists, however, as Watson and his followers have made us realize, such a conception explains nothing and is entirely redundant. Professor Burt apparently believes that he is able to identify some such entity by 'introspection', i.e. by direct intuitive self-observation. That merely illustrates how unreliable introspection can be as a scientific method. In science observation means observing things which any other investigator can also observe and check. Few psychologists claim to have observed any such thing as a mind.

Now, unlike many of the critics I have mentioned, I myself would not go so far as Watson in his famous manifesto and deny the very existence of consciousness. But I should certainly agree that the recognition of something strictly private and personal, going under that name, could not possibly justify the postulation of an exclusive entity called 'mind' to account for it. The reintroduction of such a term into current psychology (as its champions frankly admit) implies the acceptance of the metaphysical concept of substance—a concept which irrefutable criticism has long since exploded. In the place of substances could we not substitute events? A flash of lightning is a brief event; a human body just a longer event; and a mountain a very prolonged event. Thus what McDougall regards as a mind or a mental structure I should regard merely as a mental process. This, if I understand them aright, is the view taken by Russell and Whitehead; and, as they point out, it effects a manifest economy in the number of basic concepts we are led to postulate. Consciousness in any case is surely not a thing or a stuff but an event or process. As such it is incessantly changing. The causes of the changes are cerebral. Certain physical events in the brain are the necessary and sufficient causes of all conscious events. This is after all the only interpretation that is consistent with the general theory of the universe adopted in modern science.

Let us recall the other mystical principle that for so long confused the lowlier biological sciences—that of 'life', to which Aristotle gave the same name, 'psyche'. Today every biologist would unhesitatingly reject the concept of a unique 'principle of vitality' or 'life', and trace the characteristics of living matter to the peculiarities of certain complex chemical molecules evolved by natural selection. In the same way, I maintain, we should, in company with most psychologists inspired by a modern biological outlook, reject the parallel concept of mind as a 'principle of consciousness', convinced that, like life, it too will turn out to be merely an emergent property of the still more complex patterns in which other chemical molecules are organized—the outcome once again of the blind processes of evolution. If a label be required for this modified type of behaviourism, I suggest the term 'epiphenomenalism' as defined (except for a slight verbal modification of my own) by Professor Drever: " the theory that the chain of causality is complete on the bodily side, so that conscious phenomena are merely the accompanying effects, and make no real difference to the overt behaviour ".

K. S. HUDSON

SIR,—The recent discussions[1] between Dr. Smythies, Dr. Levy, and Professor Burt have brought once again to the fore the problem of the philosophical dualism which psychology by its very title seems to imply. For the statistical psychologist the issue is of fundamental importance: are the factors and other variables with which he is concerned to be described in terms of their alleged conscious content (according to the practice of introspectionists like Spearman) or in terms of behaviour (as most contemporary American factorists prefer to do)? Burt evidently regards his factors as ' factors of the mind '; and in a series of articles he has reviewed the results obtained by such methods with the object of showing that what they reveal is essentially ' the structure of the mind '. The phrase has been echoed by several later statistical writers. Vernon shortly afterwards published a volume on ' The Structure of Human Abilities ' and Eysenck one on ' The Structure of Human Personality '. The altered phraseology is perhaps meant to evade any dispute over the use of the word ' mind '; nevertheless the words ' ability ' (which goes back to Galton) and ' personality ' (a favourite term with Ward) still indicate leanings towards the older introspectionist nomenclature rather than any change to a strictly behaviouristic viewpoint.

Burt and most of his students (Moursy, Banks, Howard, Gregory—to take only contributors to this journal)—in common with Thurstone in his earlier work—still cling to the term ' mind '. And yet, so far as I am aware, neither Burt nor any of the others have explained what precisely they understand by the word. In his interesting paper on ' Hebrew Psychology ' (this *Journal*, XIII, pp. 195f.) he discusses the *original* connotations of terms like ' spirit ', ' soul ' and ' mind ', which have played so prominent a part in the historical development of psychology; he himself seems still to think of mind in the old scholastic fashion as a specific kind of stuff—to borrow a favourite phrase of William James; and the description tends rather to encourage students to conceive or picture each individual as carrying round inside him a kind of supersubtle body with a structure of its own—a ' pneumatic soma ', in the Greek phrase he quotes, in short, a veritable ' ghost in the machine'. According to him, the ' mind' or ' self ', is not, as it is with Descartes, merely a simple unextended substance; it is something of which in introspection I am directly aware.

If it is argued that ' mind ' is a hypothesis, then it is, I submit, a hypothesis for which (in Laplace's phrase) there is no necessity. Thanks to advances of the basic sciences the scientific world-view has become increasingly homogeneous. Occam's razor—" entia non sunt multiplicanda "—should by now have assured us that for the scientist the universe contains physical entities and processes, and nothing but physical entities and processes. It follows that, since mind by definition is not physical, minds do not form part of the furniture of the universe. " All sentences of psychology," says Carnap, " describe physical occurrences, to wit—the bodily behaviour of human beings or other animals." Since psychologists are now generally agreed that psychology is simply ' the study of behaviour ', and since all that we observe is essentially behaviour, why cannot Professor Burt rest content to conform to the practice of his contemporaries, and describe the conclusions of factor analysis and statistical psychology *in terms of behaviour*, without dragging in a hypothetical entity like the mind, which after all is from its very nature unobservable? Can he give us any specific instances of individual differences that cannot be completely described in behaviouristic terms?

T. H. BOLTON

[1] J. R. Smythies, ' Three Classical theories of Mind ', *J. Soc. psych. Res.*, *XL*, 1960, pp. 385f.; cf. Burt, *ibid.*, *XLI*, 1961, pp. 55f.; and Levy and Burt, this *Journal*, *XIV*, 1961, pp. 51f.

THE STRUCTURE OF THE MIND

A Reply

By Cyril Burt

I. Differences in Methodology

The Definition of Mind. The primary object of this paper is to plead for a reintroduction into psychology of the concept of mind. In support of that plea, I shall endeavour to explain in fuller detail what I take to be the nature of mind, and at the same time try to answer my various critics. Hitherto, when summarizing the conclusions reached by researches on statistical psychology—for instance, in *The Factors of Mind* and articles on 'The Structure of the Mind'—I have used the word without stopping to define what precisely I mean by it. In an earlier work[1] I briefly mentioned that I "used the term as a convenient label to designate the hypothetical basis of consciousness, whatever it may be"; and added that we ought rather to refrain from deciding in advance whether it was or was not material, or spatially extended, or even identifiable with certain parts or tissues of the brain. It seemed to me that any detailed description of the mind was a matter to be settled not by definition at the outset, but by deductions step by step from the conclusions reached as our investigations advance.

The Principle of Parsimony. The question with which most of my critics begin is whether the assumption of an immaterial mind is 'really necessary'. Dr. Bolton, for example, inquires "why, since psychologists are generally agreed that psychology is the study of behaviour, and since all that we can observe is essentially behaviour, Professor Burt cannot rest content to adopt the practice of his contemporaries, and express his conclusions solely in terms of behaviour, without dragging in a hypothetical entity like the 'mind', which from its very nature is unobservable?" My answer is that nowadays few scientists outside the ranks of psychology accept the methodology by which Dr. Bolton supports his arguments—that of the operationist, the logical positivist, and the devotees of Occam's razor. Certainly none would dream of banning a useful theoretical concept just because it is unobservable. After all, even in the material sciences, what the physicist and the chemist actually observe is behaviour. But they do not (as Dr. Bolton demands) "rest content with expressing their conclusions solely in terms of behaviour". No one could 'observe' the nucleus of an atom; yet that did not stop Rutherford from postulating such an entity to account for the behaviour of alpha-particles when fired at a thin film of gold or silver foil; nor did it prevent later physicists from depicting the unobservable nucleus as an aggregate of just so many particles (or rather quasi-particles) and endowing it with an entirely new kind of 'binding energy'. I claim a similar freedom for the psychologist.

Dr. Bolton, however, cites Carnap, and contends with him that "Occam's postulate should have convinced us that, for science at any rate, the universe contains physical and nothing but physical entities; it follows" (he continues) "that, since mind is not a physical entity, minds are not part of the furniture of the universe". It is precisely at this point that our views diverge. If he and other behaviourists and physicalists were content to apply Occam's razor simply as a working principle, I should readily agree; but, when they tell us to accept what it dictates, not as a working principle, but as an empirical premiss, I demur. In any case, so far from becoming "increasingly simple and increasingly homogeneous", as nineteenth century physicists and Dr. Bolton have maintained, the "scientific world-view" has, during the last sixty years, become increasingly complicated.

---

[1] *How the Mind Works* (1933), pp. 35f., cf. pp. 331f.

But what I chiefly wished to protest against in the articles which my various critics cite was the behaviourist's regular practice of regarding certain types of elusive or anomalous problems as already solved by direct deduction from their 'homogeneous world-view', and dogmatically denying the need for any *ad hoc* verification of the facts they allege. Mr. Hudson rejects all problems requiring introspective study. Mr. Wilson rejects any theory or reported observation implying the efficacy of will. Mr. Hansel assures us that it is a waste of time to seek or sift evidence for telepathy because the laws of mechanics preclude any such thing. Dr. Moseley, and, if I understand him rightly, Mr. Gregson, "reject the notion that in the course of conscious experience we may directly apprehend certain objective qualities called values", because value can not possibly form an objective element in a physical universe governed solely by physical laws. And in much the same spirit Dr. Taylor and Dr. Levy recommend the psychologist to "keep well within the bounds of Newtonian science, since that is the surest way to avoid the invalid postulation of subjective qualities or occult entities, like consciousness and mind".

But science never attains finality. As its history shows, whenever men have cited the current scientific picture of the universe as sufficient in itself to dispose of any novel theory or unexpected observations, they have nearly always been wrong. Consider how, right down to the eighteenth century, the occasional cranks who believed that the singular properties of magnets and of amber might be worthy of study were persistently ridiculed as "credulous time-wasters". Consider how even Newton's theories were challenged because they "revived the occult doctrine of action at a distance, which is no better than an outworn superstition".[1]

The Need for a Concept of Mind. In citing Occam's razor—'Entia non sunt multiplicanda'—Dr. Bolton omits the final phrase 'praeter necessitatem'; and it is my contention that there is an inescapable necessity for the concept of mind. One undeniable fact has to be faced by every psychologist, though behaviourists seem almost wilfully to close their eyes to it. In our everyday conscious life each of us is indisputably aware of certain phenomena—our experiences of colour, sound, warmth, pain, pleasure, and its opposite—which cannot possibly be reduced to terms of purely physical or mechanical processes; hence physicalism or mechanism cannot possibly be the *whole* truth about the universe. And, even if we grant that such experiences are mere 'epiphenomena'—transient appearances produced by certain specifiable physical processes, nevertheless the laws or principles relating the physical processes to the psychical experiences cannot themselves be physical laws.

My chief concern therefore has been to affirm the need for some hypothetical framework which, unlike that of the prevailing behaviouristic schools, should not rule out all reference to the contents or the processes of *consciousness* from the very start. For example, in the articles on factor analysis which my critics cite,[2] the test results there summarized seemed to indicate that individuals differ significantly in a wide variety of capacities which can only be described in terms of conscious experience, e.g. capacities for understanding meaning, relations, values, for generalizing, for reasoning, etc., etc. Not one of these can adequately be described in purely behaviouristic language, or in the anatomical, physiological, and biochemical terminology employed by the neurologist;[1] and attempts to treat colour-blindness, eidetic images and the like, as no more than modes of behaviour have resulted in a very misleading presentation of the facts.[2]

---

[1] Leibniz, for example, wrote: "That one body should attract another with no intermediary means is not so much a miracle as a sheer contradiction; for 'tis to suppose the body can act in a place where it is not" (*The Leibniz-Clarke Correspondence*, ed. H. G. Alexander, 1956, p. 72). Locke similarly contended that "it is impossible to conceive that a body should operate on what it does not touch" (*Essay concerning Human Understanding*, II, viii, 11: cf. Aristotle, *Physics*, VII, 242 *b*). In a later edition of his *Optics* Newton endeavoured to circumvent such criticisms by declaring: "These Principles" (gravity, etc.) "I consider not as Occult Qualities . . . but as general Laws of Nature" (*Query* 31).

[2] 'The Structure of the Mind: A Review of the Results of Factor Analysis' I & II, *Brit. J. Educ. Psychol.*, XIX, pp. 100–111 and 176–199.

The Mind as a Model. Mr. Hudson, however, seems quite to misunderstand my position when he suggests that I have put forward my concept of mind as " an empirically established theory ", and asks " what exactly is the observational evidence on which [my] proof is presumed to rest ". All I sought to outline was a very tentative and flexible concept, put forward as a provisional working model rather than a descriptive ' theory '. As to the precise nature of the model proposed there seems to be a considerable degree of misunderstanding. The views that Dr. Bolton and Mr. Hudson attribute to me are views that I should repudiate as vigorously as they do. Mr. Hudson regrets that I should be " resurrecting the traditional Cartesian notion of a simple unextended substance, called a ' mind ' or ' soul ' "; and Dr. Bolton, recognizing that in an earlier paper I had expressly said that I did not regard the mind as " necessarily unextended ", complains that I am " encouraging people to conceive or picture each man as carrying around with him a kind of super-subtle body with a structure of its own—literally a ghost in the machine ". I realize that a word like ' mind ' is apt to call up irrelevant associations of this kind. But in point of fact the type of entity I had in view was neither a ' super-subtle body ' nor a ' simple unextended substance ', but something very different—namely, a ' field '.

And if I am asked—" Why have a model at all? Why not be content with a purely formal statement of your hypothesis?"—then I answer that the use of interpretative models has become part of the practice—and a very fruitful practice—of every modern scientist. The Rutherford-Bohr picture of the atom, with its heavy central nucleus and lighter electrons revolving round it like planets round a sun, was a typical scientific model; like all such models it was avowedly put forward as an abstraction capable of a variety of modifications, and in the end was modified almost out of existence. According to Lotze, " every successful hypothesis is based on an analogy ". The analogue or model supplies the puzzled investigator with a familiar system whose structure and working are already well understood. And in using analogues to construct our tentative models we rely on what Keynes[3] calls the ' known positive analogy ', and admit the existence of a ' known ' (and in all probability an ' unknown ') ' negative analogy ', i.e. features in the analogue which are inapplicable and which we therefore treat as inessential. The quantum physicists' ' particles ', for example, are not intended to be taken literally, since they are quite unlike any particles we encounter in our everyday experience. The physicist is even prepared to adopt several different models for the same situation according to the particular aspect or question that interests him at the moment. The behaviour of any atomic nucleus which contains more than two neutrons presents problems too complex to be solved by direct attack; so the physicist falls back on various models none of which forms a perfect analogue—the liquid drop model, the Fermi gas model, the nuclear shell model, and so forth. And, since present day psychology seems in much the same confused condition, I see no reason why psychologists likewise should not exploit a number of different models simultaneously instead of quarrelling about the merits of different schools, or propounding yet another school " to end all schools ". For certain purposes I should be quite satisfied to adopt the behaviourist's model; for others I should feel obliged to construct a model of my own—yet even that model, I suspect, would prove to be something of a composite.

---

[1] Dr. Bolton asks for " specific instances of individual peculiarities that cannot be completely described in behaviouristic terms ". I have space only for one or two from my own personal experience. It so happens that my right eye perceives colours rather differently from my left (they have a somewhat stronger tinge of yellow or brown); Dr. Bolton could not possibly infer this peculiarity from my behaviour. Or again, two of my weakest capacities are spatial comprehension and mechanical retentiveness: I find no one ever seems to guess this from my actual behaviour. In answering tests for such capacities I quickly translate the problems into forms more suited to my personal habits and experience, and thus obtain marks which are deceptively high. And the same is true of certain types of deductive reasoning, where I often get more easily confused than most people unless I mentally change the logical scheme.

[2] E.g., J. L. Fuller and W. R. Thompson (*Behavior Genetics*, 1960) have endeavoured to treat all problems of mental inheritance as problems in the inheritance of behaviour.

[3] J. M. Keynes, *A Treatise on Probability*, ch. xix, ' The Nature of Argument by Analogy '.

II. Current Views

Mind as Substance. Dr. Hudson criticizes " the re-introduction of ' mind ' into psychology " because, he says, " the notion of a ' mind ' implies the acceptance of the metaphysical concept of substance—a concept which irrefutable criticism has long since exploded ". The word ' substance ', however, has been used in a variety of ' metaphysical ' senses; and Dr. Hudson fails to explain which particular meaning he himself attaches to the term. From his other remarks I suspect that what he is thinking of is, not so much the more technical usages to be found in the writings of the philosophers or scientists, but rather the narrower sense which the word so often bears in popular speech—" the tangible material out of which a commodity is made ", to quote the dictionary definition.[1]

Explicitly or implicitly all the natural sciences use the traditional concept of substance, even if they do not employ the word; and they use it chiefly as a matter of convenience, without committing themselves to any specific metaphysical doctrine. My own usage was much the same. However, since Dr. Hudson has now thrown down a challenge, a fuller explanation seems needed.

To avoid the misunderstandings which inevitably arise when one introduces popular terms—with their trail of popular associations—into scientific expositions, it will be necessary to fall back on a somewhat esoteric terminology, so that words employed can be unambiguously defined. Let us begin by distinguishing, within the sum total of reality, two broad categories: (i) " particulars, which enter into complexes only as the subjects of predicates or as the terms of relations, and which, if they belong to the world of which we have experience, also exist in time "; (ii) " universals, which occur either as predicates "

---

[1] In its more precise sense the word ' substance ' (a Latin word coined by analogy with the Greek) is a term of logical exposition rather than of metaphysics. It goes back to Aristotle, who interprets the concepts of ' substance ' and ' attribute ' by reference to the logical notions of ' subject ' and ' predicate ': " a primary" (i.e. concrete or individual) "substance", he says, " is that which is neither asserted of a subject nor present in a subject" (*Categ.*, 5, 2a, 11). The later ' metaphysical concept ' is traditionally defined by saying that a substance is *res per se subsistens*, i.e. " an individual self-existent entity characterized by a certain degree of unity and of permanence in time, and capable of standing in causal relations ". McTaggart, who criticizes the notions of both cause and time, suggests an alternative formula: " a substance is that which has qualities and is related, without being itself either a quality or a relation or having qualities or relations among its parts " (the last clause is added to exclude ' facts '). His wider phraseology is intended to take in certain relatively transient ' particulars ', which many philosophers would not regard as substances, e.g., a sense datum (J. E. M. McTaggart, *The Nature of Existence*, I, p. 68; cf. *id.*, *Contemporary British Idealism*, I, p. 253). The usual objection (which presumably Dr. Hudson has in mind) is that, as traditionally interpreted, substance turns out to be a mere peg on which to hang qualities or attributes, just a " something I know not what ". Many writers treat the word ' substance ' as synonymous with ' thing '. But a ' thing ' is usually considered to be not only persistent in time, but also bounded in space—a condition which can only be fulfilled at temperatures similar to those ordinarily obtaining on the earth: in the centre of the sun ' things ' (as thus defined) would be impossible. Moreover this restriction would forbid our conceiving either energy or Cartesian mind as ' substance '.

(i.e., as " attributes ") " or as relations, and do not exist in time ".[1] The terminology and the definitions are Russell's. However, as Johnson has pointed out,[2] the word ' particular ' may give rise to awkward side-issues: for, when a quality or relation is predicated of a ' particular ' subject, it is tempting to argue that the quality or relation must also be in some sense a ' particular '. I shall therefore adopt the alternative name ' existent '; and I shall follow Johnson in assuming that " the category of the ' existent ' may be divided into two subcategories: (i) what continues to exist or the ' continuant ', and (ii) what ceases to exist or the ' occurrent ', every occurrent being referable to a continuant ". " A ' continuant ', " as he says elsewhere, " continues to exist and remains identical with itself, although its states or relations may be changing." His instances include the psychological as well as the physical. " The body which illuminates the earth is identical with the body that warms it," and " the person who was experiencing the toothache is identical with the person who now intends to go to the dentist."

The statement, therefore, that mind is a ' substance ' need mean no more than this: first, that mind is not a ' universal ' (i.e. not a quality or a relation, as consciousness is often held to be), but rather a ' particular ', or, as I should prefer to say, an ' existent '; secondly, that it is not an ' occurrent ' or a mere ' event ', but a ' continuant ', i.e., an existent that evinces some degree of persistence in time: by this I mean that, although its states and relations may be constantly changing, the modes of change consistently exhibit some kind of regularity and law, such as is usually described by assigning so-called ' dispositional properties ' to an underlying ' existent '.

Mind as Process. Several writers have attempted to dispense altogether with the traditional concept of ' substance ', and have substituted certain types of ' occurrents ', which they term ' events ' or ' processes '. As Dr. Hudson implies, this effects a manifest economy in the number of basic concepts that we postulate to account for the universe as we know it; and, largely for this reason, Russell in his later writings " treats as fundamental the notion of ' events '." By an ' event ' he means " something occupying a finite amount of space-time ", e.g., " seeing a flash of lightning ". Events may have parts; but minimal events have no parts.[3] Things (i.e. pieces of matter), and minds, and indeed " everything in the world ", are ' constructed out of events '. Broad prefers to talk of ' processes '. His favourite example is a noise, which he holds may be adequately described without reference to anything causing the noise or anyone hearing the noise. And he argues at some length that processes may be ' absolute ', i.e., may exist without relation to any substance which may cause them or in which they may occur.

In that case, if we use the word ' mind ' at all, it would designate, not a persistent underlying substance or structure, but the process or processes making up an individual's conscious life-history. On this basis Broad has tentatively developed a form of epiphenomenalism. He holds that " we can take the notion of ' mental event ' as fundamental ", and replace the traditional notion of a ' mental entity ' by something consisting solely of mental events and the relations between them.[1] This doctrine goes back to Hume, who in a well known passage declared that " they are the successive perceptions only that constitute the mind. There is no simplicity in it at one time nor identity in different. I venture to affirm that it is nothing but a bundle or collection of different perceptions. I never catch *myself* at any time without a perception, and never can observe anything but the perception ".[2] Much the same doctrine seems still to be regarded as the orthodox psychological theory of the day by many physiologists, by most medical men, and by those psychologists who still use such terms as ' mental ' and ' conscious '. And in the same spirit Dr. Hudson argues that " what McDougall regards as a mental structure [he, Hudson] would regard merely as a mental process ".

---

[1] B. Russell, ' On the Relations of Universals and Particulars ', *Proc. Arist. Soc.*, XII, 1912, pp. 23-24; cf. also *id.*, *Problems of Philosophy*, chaps. ix and x: here we are told that " Particulars *exist* and Universals *subsist* " (a phraseology borrowed from Meinong). In his later writings, however, Russell argues that " there is no need to assume particulars (usually regarded as located in time and space) as subjects in which qualities inhere " and that in their place we need only postulate " bundles of qualities ". But to my mind he has not succeeded in overcoming the difficulties which this alternative formulation involves—difficulties which he himself had pointed out in the article just cited.

[2] W. E. Johnson, *Logic*, Pt. II, p. xi; cf. Pt. I, p. 199. The word ' existent ' may itself be a little misleading. Although it has the form of an adjective, it must not be taken as designating an attribute. I myself should hold that ' existence ' cannot be defined, though it may be indicated.

[3] *Outline of Philosophy* (1927), pp. 287f. This view, as he explains, was originally due to a discussion with Whitehead. In his later writings, however, he tells us that "there is a possibility of a further stage of analysis in which events are no longer the *ultimate* raw material ". With the theory as thus qualified I myself should largely agree, though my " further analysis " would probably be different from Russell's.

The Inadequacy of the Process Theory. It was this view that I sought to criticize in the discussions to which Mr. Hudson refers. For certain limited purposes—particularly those of the medical psychologists and the physiological research-worker—it may no doubt furnish a rough and ready working model. But for the student of general psychology, concerned specifically with the fundamental features of our conscious life and its underlying laws, it is in my view not only grossly inadequate but seriously misleading.

To begin with, it is difficult to think of any convincing instance of a process existing in its own right. If noise means the sound I hear, it ceases to exist when I stop my ears; if it means the physical stimulus, that exists only as a vibration of the atmosphere. Thus noise, and, so far as I can discover, all other processes appear merely to be constituents in a composite situation which contains one or more substances as their basis; and, if our theorists reply that these substances can in turn be reduced to bigger processes, then, simply from a blind belief in parsimony, we seem to be abandoning a most useful distinction and removing ourselves further away from the facts as we actually observe them. Secondly, a substance, as we have seen, has a certain degree of unity and permanence. The silver medal which I saw in my drawer this morning is the same as that which I received as a schoolboy sixty years ago: it has changed a little; it is somewhat tarnished and slightly worn. But if the medal I received was just a ' process ' then, even supposing the ' process ' that occurred sixty years ago appeared almost identical with the ' process ' that occurred this morning, I still could not call it the *same* process. The slow *changes* that have taken place in the meantime no doubt form a ' process ', and the process may claim some degree of continuity. But the word change is meaningless unless there is a relative permanent something which changes. Underlying the manifest changes, science seeks—or at any rate postulates—some kind of latent *invariant* in terms of which to formulate the regularities or ' laws ' that govern the changes. Thirdly, most philosophers, I fancy, and certainly Broad, hold that processes themselves possess no dispositional properties.[3] In a textbook on chemistry, for example, we find enumerated under the heading of each element its physical and chemical ' properties ': we are told that copper will melt at 1084° and boil at 2310°; that, like other metals, it will conduct heat and electricity; and so on. These properties are regarded as in some sense inherent in substances; and it would seem very far-fetched to treat all such statements as really referring simply to ' processes ' that take place in the void.

---

[1] Cf. *The Mind and Its Place in Nature* (1947), pp. 587f.

[2] *A Treatise of Human Nature*, Bk. 1, pt. iv, ' Of Personal Identity '.

[3] It might in certain cases be possible to modify the ordinary conception of ' events ' or ' processes ' so that they did possess dispositional properties. Mr. Hudson's illustrations—" an explosion is a brief event rather than a substance, a human body is just a longer event, and a mountain a *very* prolonged event "—suggests that he has something of this kind in mind. But the final result would be, not to eliminate the concept of substance, as I have defined it, but simply to give it a new name: instead of a ' continuant ' it becomes a continuous ' event ' (a phrase which is almost a contradiction in terms). And the whole contention turns into a purely verbal issue.

But in point of fact hardly any scientists—not even the physiologists and psychologists who want to reduce *mental* substances to ' processes '—seek to apply the same kind of reduction to *material* substances. The neural processes which condition mental phenomena are almost invariably regarded as processes taking place within a material structure or substance, namely, the brain.

Epiphenomenalism. If, with the physiological materialists, we treat the brain and its various properties as fundamental and the processes of consciousness as a mere stream of transient events, resulting from ' changes in the brain-substance ', then we are naturally led to some type of ' epiphenomenalism '. In its commonest form the doctrine of epiphenomenalism may be summed up in terms of two simple propositions. (1) Physical events in the brain produce conscious events in accordance with certain definite causal laws, some of which have already been discovered and well established; but (2) conscious events play no part in the production either of (*a*) physical events or (*b*) other conscious events. The causative relations between brain and consciousness are thus strictly unilateral.

Now it is not difficult to show that the adoption of a one-sided theory of this kind entails several consequences which are unquestionably paradoxical if not mutually incompatible.

(1) Let us begin with the first of the two propositions that make up the theory. It implies that certain physical processes, taking place in a piece of matter called the brain, constitute the necessary and sufficient causes of all conscious processes. This view is put forward as " the only explanation which is consistent with the general theory of the universe adopted by modern science ". But almost at the outset it issues in a flagrant inconsistency. One of the basic principles of modern science is the so-called law of the conservation of energy. Now in producing changes of consciousness the physical changes in the brain must either expend energy or not expend it. If they expend it in this particular way, then that amount of energy must be completely lost, since conscious changes can never restore the energy, for by hypothesis they have no influence on physical processes. If, on the other hand, no energy is expended, we are faced with the paradoxical phenomenon of observable changes being produced without energy being needed for their production. Such a phenomenon is found in no other field of science.

But let us waive this familiar objection as a mere *argumentum ad hominem*, and consider the proposition on its merits. In every other branch of science, when a causal process is postulated, there is always some degree of generic likeness (though not perhaps of specific likeness) between the cause and its effect. The most perfect instances of causal explanation are those cases in which mechanical causes produce mechanical effects, as when one billiard ball imparts its motion to another. Even those who question whether causation demands *complete* resemblance between effect and cause nevertheless accept the principle of continuity in some weaker form. They would argue that, if the *whole* cause is not of the same generic type as the effect, at least part of it must be: from which it would follow that at least part of the cause of a mental effect must be mental, or (to put it in another way which some would prefer) the material change provides the occasion for the mental change rather than the complete cause. If material conditions together with certain material changes formed the sole cause of a conscious experience, then this would be ' creation ' not ' production '. It would be, as Stout puts it, " the miracle of making something out of nothing ".[1]

To overcome this paradox the materialist commonly tries to suggest that consciousness is in some way unreal, or even to deny it any sort of existence. Hobbes long ago referred to conscious occurrences as mere ' apparitions ' or ' phantasms '—a metaphor echoed by Professor Ryle when he jeers at the ' ghost in the machine '. Huxley and the reflexologists brush them aside with a string of mixed metaphors: they are " the mere collateral by-products of the proper functioning of the brain . . . a foam, a froth, a phosphorescent emanation, the noise ground out by the neural machinery in the course of its working, the shadow thrown by the train as it moves on its journey "—in short " epi-phenomena " in the literal sense—" mere superfluous appearances ". And yet if the pronouncements of our consciousness—our first-hand observations and the rational inferences based upon them—are all unreal, what grounds have we for asserting that anything is real?

---

[1] For a fuller exposition of the argument see G. F. Stout, *Mind and Matter*, pp. 116f. and J. Wisdom, *Problems of Mind and Matter*, pp. 83f.

(2) The second proposition involved in the epiphenomenalist hypothesis is that conscious events, though caused by physical events, do not themselves act as causes: they produce neither physical changes nor changes in other conscious processes. This again is a denial of what seems so obvious to common sense—a denial made solely to preserve the oversimplified picture of the universe that is supposed to be part of ' modern science '. And once again in belittling consciousness the critic is belittling the source of his own concepts. For where after all do we get our notion of causation, if not from the direct experience we seem to have of causation in conscious volition? When we ' will ' make to a movement of our own, we seem to be immediately aware of a specific tendency issuing in a corresponding effect. By analogy we infer a similar process of causation when one material object appears to produce movement in another. The stimulation of the sense-organs of the body is invariably followed by corresponding sensations; therefore, says the epiphenomenalist, the sensations must be caused by the stimulation. But in precisely the same way the conscious decision to make this or that movement is invariably followed by stimulation of the corresponding bodily muscles. Had we no preconceived theory to support we should unhesitatingly infer that the muscular contractions must be produced by our voluntary efforts. The natural inference is rejected only because it is assumed that a genuinely scientific account of the universe must depict it as subject to a rigorous causal determination which leaves no room for anything like freedom of the will. But now that the majority of physicists have discarded the nineteenth century notion of rigorous causal determination the main ground for denying the natural inference has disappeared.

The materialist, however, not only contends that conscious processes have no effect on bodily processes; he also insists that they have no influence on subsequent conscious processes. This again flies in the face of what seems to be common sense. If we can trust the prima facie evidence of our everyday experience, we constantly encounter cases in which one conscious event produces or affects another. And indeed we often seem directly to see not merely *that* the conscious consequents follow from the conscious antecedents, but even *how* they follow; and we see it far more clearly than we see how one mechanical change follows from another. My consciousness of the premisses of a deductive argument, for instance, appears definitely to cause me to draw, consciously and explicitly, the logical conclusion; and social intercourse would become quite impossible unless we assumed that the perceptions, beliefs, and interests of experiencing individuals are directly operative as effective conditions determining the subsequent course of their thoughts, emotions, and mental efforts.

Moreover, the mode of determination is itself peculiarly mental. The whole organization of the successive states and phases in my individual consciousness seems to imply a specifically mental type of relation quite different from the types of relation that obtain between mere[1] physical events. Both might be regarded as species of causal determination in a broader sense; but if so, we must recognize an inter-mental type of causation as well as an inter-physical. The fact that this assumption, or rather the hypothetical working models of the mind based on this assumption, serve a useful pragmatic purpose in guiding both theoretical research and practical treatment is shown by the successes of psychoanalytic theory. In the past most psychologists took it for granted that one conscious process determined another. But, as Professor Mace has pointed out, few have turned this ostensible fact into an explanatory principle; " Freud seems to have been almost the first to take *mental* determinism seriously as a basic explanation in psychology ".

---

[1] I insert the word ' mere ' to forestall objections based on superficial analogies with logical processes carried out by a ' logic machine ' or an electronic computer. In such cases of course the construction and programming of the machines themselves are the result of the designer's conscious deliberations. We are surrounded with machines and manufactured articles, which would never have come into existence unless some ingenious inventor, guided by conscious and rational steps, and prompted by non-physical qualities such as value and meaning, had desired and planned them.

It is perhaps here more than anywhere else that we come up against the most conspicuous inconsistency in the epiphenomenalist's position. He relies on conscious reasoning to prove his doctrine; and yet the very doctrine that he seeks to prove denies that conscious reasoning has any intrinsic validity. By his own showing the sole grounds for his utterances must be certain purely mechanical processes that have occurred within his brain: all else is illusion. And thus, as Stout observes, " materialism, in undermining common sense, undermines itself ".

So far the arguments I have used have been mainly negative. I wanted first to rebut the behaviourist's assumption that consciousness is at most merely a negligible by-product of the changes in some material existent—the ' organism ' or the ' brain '.[1] My next task must be to offer more positive evidence for assuming a distinct kind of existent, with characteristics of its own, entirely different from those ordinarily ascribed to material substances.

Mental Unity. To begin with, we may note that consciousness displays a peculiar kind of unity which would be quite inexplicable were it merely the effect of a miscellaneous aggregate of simultaneous and successive processes going on in a number of separate cells or synapses within the central nervous system. It presents both a simultaneous and a successive continuity.

Although at any given moment I can see a variety of distinguishable objects around me, nevertheless my visual expanse remains a single and unbroken whole: it displays nothing like the mosaic structure characteristic of the retinal rods and cones whose stimulation is supposed to generate it. And although my other sensory experiences—hearing, touch, and so forth—are disparate in quality both from each other and from my visual experiences, nevertheless they too are all fused together as constituents in a single conscious continuum. For this and similar types of continua the simplest and most appropriate name is ' field '.[2] These various sensory fields fuse into one total conscious field. This, too, is incessantly changing; and yet there is no breach, no gap, no interruption: it flows on smoothly like a never-ending stream. Even the deepest sleep or a period of complete anaesthesia causes no manifest hiatus. Often it is only after we have looked at the clock that we realize we have dropped off to sleep.

In both respects the unity of consciousness is quite unique. There is nothing precisely analogous to it in the material world. The seamless garment can be rent; the flawless globe of glass can be shattered into fragments. Consciousness can be neither cut nor

[1] Moseley cites Dr. Smythies as holding some such view. In his present note, however, (p. 141) Dr. Smythies gives up this position (if indeed he ever held it). With nearly all that he now says I find myself in almost complete agreement.

[2] What I understand by the term ' field ' has, I hope, been sufficiently explained in a previous note (' Field Theories and Statistical Psychology', *Brit. J. Statist. Psychol.*, XII, 1959, pp. 153–164). However, as Mr. Foster rightly observes, I there confined myself to explaining the concept " without giving reasons for making it an essential part of psychological theory ". Mr. Foster is wrong in supposing that " in physics a field always presupposes a material object that produces it—a magnet, a rod carrying an electric charge, a wire carrying an electric current, a body like the earth attracting other bodies in virtue of its mass ". In the theory of relativity it is, so to speak, the structure of the field that is assumed to produce the phenomenon of a body with the power of attraction rather than the material body that is assumed to produce the gravitational or magnetic field. Whether " a psychic field could exist without a brain " I am not prepared to say: there is little evidence to support such an assumption, but I know of no evidence which positively excludes it.

split. And, what is still more curious, the consciousness of one man, though it constitutes a uniform whole within itself, is completely separate and insulated from the consciousness of every other. In this twofold fashion consciousness seems to be essentially *personal*.

No doubt we sometimes talk of 'mental events' or 'conscious events', as if they were in some way similar or parallel to distinct 'physical events'. But the plural is misleading. A man's conscious life forms just one continuous event, beginning at birth or a little before, and lasting at least until his body dies. This unity and continuity strongly suggests (as indeed common language almost universally assumes) that the constituent events are related to some permanent and central entity, an entity of a special non-material kind, in short, to a personal self, who, so to speak, owns these events, and refers to them as "*my* conscious experiences or states", and describes himself by the proper name of 'I'. The only plausible alternative would be to suppose that they cohere into an integral system by virtue of their own intrinsic relations to each other—an alternative which seems rather to restate the situation than to explain it. Now I fancy most introspectionists would agree with Hume that we never *directly* apprehend this central entity or 'personal self', though many, if not most (as we shall see in a moment) get the impression that there is such an entity somehow implicitly involved in the total situation. Yet, even if we dismiss such vague impressions—which admittedly carry little weight of themselves—it is difficult, as Broad remarks, "to imagine any reason why what is a pure result of system should appear to have a unity of centre. That the mind *does* appear to be of this latter kind", he continues, "seems pretty certain; and this must be regarded *pro tanto* as favouring Central Theories".[1]

The epiphenomenalist is forced to account for the unity and continuity of consciousness in terms of the unity and continuity of the material structure of the brain. That is rather like explaining the unity of a symphony concert by the unity of the concert hall in which you hear it. In any case on this anatomical hypothesis it would follow that, if in any individual the living brain could be completely divided into two or more parts, his consciousness would similarly be divided into the same number of parts. McDougall suggested a crucial experiment. "Should I ever be seized by an incurable illness," he announced, "I have asked Sir Charles Sherrington to cut through my corpus callosum with his scalpel, and so separate my two cerebral hemispheres: I hope I shall then be able to say whether I am still a single personality or have become split into two conscious streams."[2]

Surprisingly enough, such experiments have since been carried out, though not on McDougall. "Perhaps the most astonishing of all observations in the sphere of cerebral surgery," writes Dr. Wilkie, "are those made on human beings in whom the whole corpus callosum and anterior commissure have been severed, thus cutting all known direct communications between the left and right hemispheres." One or other of two conceivable consequences were anticipated as a result: "either the patient would report distinct but concurrent activities due to the distinct but concurrent processes occurring in the two hemispheres, or one hemisphere would dominate, while the activities of the other would be suppressed". Contrary to expectation, neither of these after-effects is reported. Types of activity requiring the cooperation of both hemispheres are still observed. "Binocular fusion and the binocular perception of depth," we are told, "are not materially altered . . . The patient suffers almost no inconvenience (*sic!*) . . . Changes in his behaviour can be detected only by refined techniques; and these turn out to be rather trivial, consisting chiefly in a lengthening of reaction time."[1]

---

[1] *The Mind and its Place in Nature*, p. 585. Broad distinguishes two types of central theory: the 'existent' thus postulated might be "either a pure ego or a central event"—e.g., "a mass of bodily feeling". He himself, as we have seen, is "inclined slightly to prefer" the latter view.

[2] Cf. this *Journal*, IX, 1939, p. 2.

The Property of Consciousness. Just as we assume that a material substance possesses certain characteristic properties, so, I maintain, we may assume that what we provisionally call a mental substance or ' mind ' also possesses its own characteristic properties. They are essentially ' dispositional properties ', i.e., they can be formulated as general ' laws ' stating that ' *if or when* so-and-so occurs, *then* such-and-such a phenomenon may be observed '. Of these (as I have argued elsewhere) the most distinctive must clearly be the property of becoming conscious in certain particular ways under certain particular conditions.

At this point, however, my critics interpose with an alternative suggestion. " The assumption, " they say, " of a special ' psychic principle ', having the ' dispositional property of consciousness ', is quite superfluous: it is comparable to the obsolete assumption of a special ' vital principle ', having the dispositional property of life." Mr. Hudson goes on to assure us that " to-day every biologist would reject the concept of a unique ' principle of vitality ', and trace the characteristics of living matter to the peculiarities of certain complex chemical molecules evolved by natural selection: in the same way " (he maintains) " we should reject the parallel concept of mind, convinced that, like life, it will turn out to be merely an emergent property of the still more complex patterns in which other chemical molecules are organized—the outcome once again of the blind processes of evolution ".

My answer is that the two assumptions are by no means ' comparable ' or ' parallel '. I find no difficulty in imagining that, at a particular stage of the earth's history, in an atmosphere containing water vapour, methane, ammonia, and the like, and as an effect of strong ultra-violet radiation or of radiation arriving in the form of cosmic rays, certain amino acids might first be formed, including perhaps simple catalysts and autocatalytic products, and that these in turn might by gradual stages combine to produce polypeptide chains of increasing complexity, and eventually throw up macromolecules of deoxyribonucleic acid, possessing the property of self-reduplication and the power to control enzyme synthesis. But Mr. Hudson's version overlooks one important fact: somehow or other the molecules have to get localized and organized into minute and inconceivably intricate structures known as ' cells '. That *this* should be the outcome of the ' blind process of evolution ' (and all within the short space of about one or two thousand million years) I find it far more difficult to imagine. Nor is there anything like the unanimity of view that Mr. Hudson supposes. Walter Elsasser, for example, has argued (in *The Physical Foundations of Biology*) that the facts compel us to postulate certain ' biotonic laws ' (special laws of ' emergence ' I suppose Mr. Hudson would term them) which are incapable of being reduced to the laws of physics and chemistry. And Professor Eugene Wigner has endeavoured to calculate the probability that self-productive systems could come into existence as an automatic result of purely blind or random processes: he claims to show, not merely that the probability is exceedingly small, but that it is exactly zero.

Such arguments, however, can hardly be regarded as conclusive. So let us provisionally admit that what Mr. Hudson says about ' life ' may be true. His conclusion could still have little bearing on the problem of ' mind '. His inference really turns on the ambiguity of the term ' emergence '. The biological materialists to whom he refers see no need for such a word. Given certain conditions at an early stage of the planet's history, then (they maintain) complex substances having the properties of growth and reproduction would in the end automatically be formed, and what we take to be the distinctive characteristics of living matter would in principle (despite Elsasser's contentions) be deducible from the laws of physics and chemistry.[1] Nevertheless, it remains utterly inconceivable that *consciousness* (in the sense of direct awareness) could possibly be deducible from those laws. Thus in the one case 'emergence' means 'is in the end automatically formed'; in the other it means 'arises unexpectedly as an entirely new property'. But to stop with the blissful word 'emergence' is not of itself an explanation. Mr. Hudson's simple answer to the problem of origins resembles Topsy's theory of her own appearance in the world: "'Specs I just growed". In all such cases (as the history of electromagnetism and quantum theory sufficiently shows) some ardent scientist would at once start out to discover what fresh concepts or principles were needed to describe the 'emergence' of these novel phenomena.

---

[1] For a more detailed summary of the results, with references to the neurological literature, see J. S. Wilkie, *The Science of Mind and Brain*, 1953, pp. 133f. For reasons that would be here out of place, I should not myself regard such observations as decisive in themselves. In psychology no single type of experiment, however impressive, can be absolutely conclusive, since alternative interpretations are almost always conceivable.

Every conscious process or state involves as its essential feature the relation of awareness.[2] But a relation can only hold between two or more terms. Thus in perception there must be a *perceiver* who is capable of perceiving as well as a *percept* which is actually perceived. And I maintain that this is a description of fact, not a mere grammatical convention. The conclusion to be drawn, however, is not that the perceiving is done by a 'Pure Ego'. The 'Ego' that I myself should infer would be highly impure. Indeed, so far as the present argument takes us, I see no reason why the perceiver should not be the brain or some particular type of tissue within the brain. But in that case the brain or the tissue would be very much more than a mere material substance or a mere succession of physical events: it would itself possess the characteristic properties of what most people would call a 'mind'.

Finally, I hold (as I have argued elsewhere) that in introspection I may become directly aware of this awareness. I perceive that I am perceiving. And here, as we have already seen, the impression that most impartial observers get is that in their own case the perceiving seems to be done by something which has the qualities, not of a palpable substance like the brain, but rather of what the psychologist commonly refers to as a 'self'. Now, when I make a statement such as: "I notice (or 'I judge') that I am undergoing this or that experience", the subject—'I'—is mentioned twice; and the form of statement implies that the subject who makes the judgement is the same as the subject who has the experience. Or, again I may say: "I remember that last week I had much the same experience". And here also the identity of the two subjects is tacitly affirmed. The subject therefore appears to be a more or less permanent entity, namely, what I ordinarily call 'myself' or 'my mind'.

There are many other ways, more specific than these, in which mental tendencies differ from physical. I will mention only one—the manner in which conscious processes seem directly to modify one another. Our two eyes supply us with two distinct visual fields: in looking through a stereoscope we may see, first, two slightly different pictures of the same scene; then, as the eyes converge, the pictures overlap and finally coincide with a click; the moment they do so, a new sensory quality emerges—what we describe as ' depth '. This experience—the experience of directly perceiving a two-dimensional field as three-dimensional—is so anomalous and unique that it could not possibly have been predicted in advance. It is an effect that has no counterpart among effects of purely physical processes. Moreover, if we make another experiment with one field plain green and another plain blue, the resultant field has an intermediate colour resembling both, namely, greenish-blue. But if one field is green and the other red, the resultant is not greenish-red, but yellow—a colour apparently so pure that the uninstructed viewer takes yellow, and not green, to be the third of the ' pure ' or ' primary ' colours. Introduce yet another overlapping colour—add blue to the green and red (as in the familiar experiment with the triple projector)—and the result is now, not a bluish-yellow, much less a reddish-greenish-blue, but pure white. These and other properties of our mental processes differ so radically from those we discern in material substances, that it seems reasonable to postulate an entirely different type of entity underlying them. To determine in detail what are the various dispositional properties distinctive of such an entity is one of the essential tasks of psychology.

---

[1] It is true that Broad, writing in 1925, cited the example of chemical compounds, like water and sodium chloride, as instances of 'emergence'. Nothing, he argued, that we know of either element by itself would give us the least reason to expect that it would combine with the other. Today, however, our increasing knowledge of the inner structure of the atom (notably the theory that valency results from either the transference or the sharing of electrons) goes far to explain not only the phenomena of chemical composition, but also many of the resultant properties.

[2] Many psychologists have denied the suggestion that there is any such thing as a unique and irreducible 'relation of awareness'. Tacitly or explicitly they seem to hold that the relation alleged can be reduced to some other more familiar type of relation—e.g., resemblance, causation, ownership, subject-and-predicate, part-and-whole, a quasi-temporal relation ('compresence'), or (if I understand Dr. Smythies aright) a quasi-spatial relation. I have not discussed these alternative interpretations in detail in the text because I have already examined them at length elsewhere. All seem open to the following objection (as well as several others). Before we can assume that a relation, A, is reducible to some other relation R, we must first show not only (i) that such statements as $s\ A\ o$ entail $x\ R\ y$, but also (ii) that $x\ R\ y$ entails $s\ A\ o$. Those who attempt some kind of proof (or more frequently some plausible type of illustration) commonly stop short with the first of these two propositions; and further reflection almost invariably demonstrates that the second proposition does *not* hold good.

To each of the foregoing arguments, considered by itself, more or less plausible replies could no doubt be given. Taken jointly, their cumulative effect is pretty formidable. Even so, they are still by no means conclusive. Yet that could be urged against the arguments for almost every scientific generalization, and is certainly to be expected in a field where the phenomena are highly complex, and lie away from the well trodden paths followed by the older sciences. Let us therefore now inquire whether it is possible to offer some more positive or constructive theory to meet the various problems we have encountered.

III. Instrumental Theories

Earlier Versions. Among those who reject the traditional notion that the study of consciousness is a " valid and important task for the scientific psychologist ", the majority, as we have seen, have assumed that consciousness is just a collateral by-product " generated by the chemistry of the nervous tissue "—" a mere function of the brain and its workings ". The strongest evidence for such a view is to be found in the close correlation between conscious processes and brain processes. Yet the tacit assumption that the correlation between two such disparate variables is to be explained by a one-sided causal relation between them involves a fallacy with which the statistical psychologist is all too familiar. There are many alternative interpretations which will not only account for the observable correlation between cerebral and conscious processes, but also surmount many of the objections to which a purely materialist hypothesis is open. Of these perhaps the simplest is the so-called ' instrumental theory '. This assumes that the brain is not the generator of consciousness but merely the instrument by which mind or consciousness is enabled to interact with the material world.[1]

As an explicit hypothesis the idea was first mooted by F. C. S. Schiller—one of the livelier critics of the Anglo-Hegelian idealism that dominated academic philosophy in Britain towards the close of the last century. " Matter, " Schiller contends, " is not that which *produces* consciousness, but merely that which *confines its working within certain*

---

[1] It is curious that C. W. Morris in his *Six Theories of Mind* (1932) fails to include what is ordinarily known as the ' instrumental theory ' among his selection. Bergson is mentioned only as the champion of a novel brand of idealism, which borrows many of the tenets of the ' new realism '. A brief footnote refers to a phrase of his which suggests that the mind uses the body as its instrument, and argues that it would be " more consistent with his thought to say that the body uses mind as its instrument "; and it is in this sense that Morris later discusses what he calls an ' instrumental theory '.

limits . . . Material organization does not manufacture consciousness out of arrangements of atoms, but contracts its manifestations within the sphere which the particular organization of matter permits." Later the theory was taken up and amplified by William James in America and by Henri Bergson in France; and it is in these more elaborate forms that it has usually been discussed and criticized.

The Brain as the Instrument of Consciousness. James assumes that consciousness is a kind of stuff or substance, capable of being divided and reunited like lumps of putty or drops of oil. It exists independently of any material organism, possibly (he says) in the form of " disseminated particles ", which may be temporarily combined into personal form by the brain, but more probably in a " vaster unity " as an " ocean " or " reservoir of consciousness " which can be temporarily subdivided into personal streams. Bergson uses much the same metaphors. Consciousness he describes as " an immense impersonal sea ". Matter in general is " an impeder or obstructor ". But the brain permits a " current " to pass through and organizes it temporarily into a personal form.[1]

Are we then to deny the familiar statement that consciousness is a " function of the brain "? That, as James observes, depends on what is meant by function. The functions of material structures may be of more than one type: they may be (i) *productive* (as when the boiler of an engine produces steam from water); (ii) *permissive* or releasing (as when the trigger of the cross-bow releases the string, and permits the energy stored up in the bow to discharge the bolt); or (iii) *transmissive* (as when the keys and stops of an organ transmit the air to the pipes). When physiologists describe ' thought ' (i.e. consciousness) as " a function of the brain ", they tacitly assume, says James, that it is a function in the first of these senses: the material structures of the brain are supposed to ' create or engender ' certain effects, and consciousness is one of them. " My own thesis," he writes, " is this: when we talk about ' Thought ' as a function of the brain, we are entitled to consider *permissive or transmissive functions.* " According to his view therefore the function of the material brain is to transmit limited amounts of consciousness from some non-material reservoir, thus permitting it to interact with the physical world. The more developed the brain, the richer will be the flow.

James, it will be seen, is here apparently adopting a modified version of transpsychism. The doctrine of *Atman* (the universal world soul) is more prevalent among Eastern philosophers and writers influenced by Eastern mysticism than among metaphysicians or psychologists in the West. During the latter half of the nineteenth century, however, it attained a popular vogue. " Die Geisterwelt ist nicht verschlossen " became a favourite aphorism. After writing his *Principles*[2] James got increasingly interested in spiritualistic phenomena, and seems to have borrowed both the conception and many of his explanatory metaphors from F. W. H. Myers' *Human Personality and its Survival.* Myers himself speaks of " our individual consciousness " as " but the phosphorescent ripple on an unbounded ocean ". According to James, what divides " the sea of consciousness " into " separate streams " is just the spatial separation of our individual brains; and only in this sense can consciousness as we know it be fairly described as a " function of the brain ".

---

[1] W. James, *Human Immortality*, 1898; H. Bergson, *Evolution Créatrice*, 1907.

[2] In the *Principles* (pp. 145–182) James has a chapter in which he severely criticizes ' The Mind-Stuff Theory ' particularly in the form of the ' mind-dust conception ' which an evolutionary psychology (he says) apparently demands. He admits that this version is " irresistibly attractive to the intellect "; yet closer examination shows that it is " both logically unintelligible and empirically false ". Empirically, direct observation should convince us that a sensory continuum, like the field of vision, cannot be compounded of separate sensations, as the ' atomistic psychologists ' had regularly assumed; it is given to us from the outset in the form of an unbroken whole. Logically the theory is unintelligible, because " all combinations that we know are *effects* wrought by the units *on some entity other than* themselves " (in the original these italicized phrases are capitalized); and this in turn implies some psychic ' medium ' on which the manifold brain-processes can combine their effects. " I confess therefore " (he concludes) " that to posit a soul seems to me the line of least logical resistance ", though to introduce such a concept into a *scientific textbook* would be to forsake the purely " positivistic and non-metaphysical attitude which empirical science should strictly preserve ".

To the theory in this form there are several obvious objections.

1. First of all, to suppose with James and Bergson that there is and can be no other form of psychic existence but consciousness as such is to abandon all possibility of devising an intelligible explanation of the peculiarities which conscious processes display. Since consciousness exists only in its overt manifestations, it has no dispositional properties. Hence, psychology, in dealing with the phenomena of consciousness, would remain purely descriptive. The only way of interpreting its special characteristics and its various changes would be to assume that they result from the dispositional properties of the brain. To do so may *look* like a reduction in the number of hypothetical assumptions; actually it increases them; for we still have to assign to the brain an assortment of unique psychical characteristics which are quite out of keeping with its physical characteristics; and, as the history of behaviourism shows only too clearly, the attempt to do so is apt to prove extremely misleading.

2. Consciousness is essentially personal and individual. It is impossible to imagine any kind of consciousness which is not somebody's consciousness, and as such it forms a completely closed and isolated system. In his *Principles* James himself placed the strongest emphasis on this unique peculiarity: " the breaches between your thoughts, my thoughts, and everyone else's thoughts are the most absolute breaches in nature: irreducible pluralism is the law ".[1] There was, however, about that time a growing tendency to believe that all forms of discontinuity were delusive appearances in a world where everything was really continuous. The atom itself was depicted as a vortex in a continuous all-embracing medium, the ether. And philosophers and psychologists alike, particularly those with idealistic leanings, were inclined to argue that the discontinuous personal form in which human consciousness appears to manifest itself is also just a superficial phenomenon—the product perhaps of similar vortices in an all-embracing ' psychic ether '. Today, with the almost universal acceptance of the quantum theory, a discontinuous or particulate mode of being no longer seems paradoxical.

3. James' notion that the division of the universal consciousness into separate individual streams might result from its temporary association with individual brains was suggested by the reports published by several contemporary psychiatrists describing how a single personality had, as a result of nervous disorder, apparently been split into two or more different personalities. Already in his *Principles* he had mentioned such cases as being difficult to reconcile with the theory of a personal soul.[2] However, except for three instances of alleged ' possession ' among mediums, all the cases cited turn out to be cases of *alternating* personality, and therefore do not really demonstrate an actual ' splitting ' of consciousness. Almost the only well authenticated instance of ' co-consciousness ' is that of ' Sally Beauchamp ', whose case was studied in great detail by Morton Prince.[3] Later observers have questioned Prince's deductions, and have completely failed to discover any similar cases, such as would confirm his conclusions.[4]

---

[1] *Principles of Psychology*, 1890, I, p. 226.

[2] *Op. cit.*, I, pp. 349, 379f. Most of the cases are taken from Janet (*L'Automatisme psychique*, 1889).

[3] Morton Prince, *The Dissociation of a Personality* (1906).

[4] Even were such cases definitely established, they could, as McDougall points out, be better explained as instances of ' possession ' by a separate pre-existent mind than by a splitting of what was originally a single personality (*Body and Mind*, pp. 370f.; cf. also T. W. Mitchell's discussion of the problem in *Medical Psychology and Psychical Research*). However, most investigators, I fancy, would today maintain that the apparent differences between the various ' personalities ' are only extreme examples of the semi-independent streams of mental activity that form a common feature of everyday life, and are phenomena of attention rather than of consciousness as such.

4. The fact is that James' theory, like all the theories postulating a psychic monism, springs from a confusion of categories. Consciousness, which should be properly described in terms of states and relations, is treated as a kind of stuff or quasi-substance. The history of science is littered with instances of this 'fallacy of reification'. Because things are spatially related, scientists have assumed the existence of a quasi-substance called 'space'; because some things are hot and others are not, they have assumed the existence of a quasi-substance called 'heat' or 'caloric'; because some things are alive and others not, they have postulated something called 'life' or 'vital force'. In the end James himself recognized the weakness of his earlier arguments; and in his later essay—'Does Consciousness Exist?'—he abandoned, once and for all, the whole notion of consciousness as a kind of stuff which could be split or compounded.

The Brain as the Instrument of Mind. Many of the objections which have proved fatal to James' version of the theory can be surmounted if we regard the brain as an instrument, not for a universal consciousness, but for an individual mind. By mind I mean, as already explained, not consciousness itself, but a hypothetical entity which is the basis of consciousness in the sense that it can stand in certain conscious relations—it can cognize, and feel, and strive (i.e. seek to attain certain purposes and goals). The simplest analogy that I can suggest to explain what I have in view is that of a machine, in the technical sense in which the word is understood by the mechanical engineer—"a systematic structure of material parts, whose relative motions are strictly constrained and limited, and by means of which the energy at our disposal is transmitted and directed so as to accomplish specific forms of work".[1] The analogy, like all such analogies, is manifestly imperfect. The brain is not merely a structural but an electrochemical machine. And I should prefer to compare the mind rather to an electrical 'field' than to a store of 'energy'. But these are side-issues. The essential assumption is that the function of the brain (or certain parts of the brain) is to release and transmit to the material world certain psychic activities of which a mind is capable. In so doing it selects those which are helpful for problems of terrestrial existence, and excludes those which are inessential for the purpose of the moment. The degree of selection and exclusion varies inversely with the complexity of the brain.

Objections to the Instrumental Theory. The most obvious objections to the theory—at least in its simpler forms—have been summarized by Professor Broad.[2]

1. His chief objection is based on reports of brain injuries and their effects on what are ordinarily supposed to be mental powers or mental characteristics. In some cases, for example, the patient seems to lose his power to remember certain events, while his other capacities remain intact. It would hardly be plausible to contend "that all such patients are lying", i.e., that they could recall the events if they wanted to; "and if", says Broad, "we reject this violent alternative,[3] we must hold that the injury does actually deprive the mind of its power to remember. Could a supporter of the instrumental theory (he asks) square such facts with his view?" In a rather different type of case the patient before the injury was "of a cheerful and benevolent disposition, and after it he became morose and liable to homicidal attacks: are we to say that the injury has made no difference to his mind?" But, as we shall see in a moment, there are many simple ways in which the instrumentalist can interpret such symptoms, without sacrificing his central hypothesis.

---

[1] A. B. W. Kennedy, *The Mechanics of Machinery*, p. 2.

[2] *The Mind and its Place in Nature* (4th ed.), 1947, pp. 533f. Cf. also *The Philosophy of C. D. Broad*, 1959, pp. 400f.

[3] There are, however, well established cases of so-called hysterical amnesia, where the patients—from shock or other emotional disturbance—are unable to recall memories; and yet (as hypnosis or free association may demonstrate) the memories have not in fact been destroyed. But in the cases referred to in the text the subsidiary symptoms were quite different from what they would be in purely functional disorders.

2. I should have thought a stronger objection could be based on the fact that direct stimulation of the sense-organs or the brain causes quasi-hallucinatory sensations—e.g., the coloured spot discernible in the visual field when one presses the side of the eyeball through the eyelid, the after-sensations produced by brilliant lights, the vivid sensory experiences and memories evoked by applying electrodes to the bared brain-cortex of patients who have not been anaesthetized. But after all, any instrument, however efficiently it works under normal conditions, may be expected to produce erratic results when the conditions are exceptional and highly abnormal.

Broad's Compound Theory. To meet these difficulties Broad himself puts forward a "modification of the instrumental theory" which he calls the 'Compound Theory'. He suggests that "what we know as a 'mind' may be a compound of two factors, neither of which separately has the properties of a mind—just as salt is a compound of two substances neither of which by itself has the properties of salt". He terms them the 'bodily factor' and the 'psychic factor' respectively; and the latter he compares to a chemical element or radical which has never been isolated. The 'psychic factor', he believes, must be capable of persisting after the 'bodily factor' has been destroyed, and of "carrying traces of experiences which happened to the 'mind' of which it was formerly a constituent"; but without the collaboration of a body it would manifest neither the higher nor the lower characteristics of a 'mind', i.e., neither thought nor sentience. "It seems a mistake," he says, "to ascribe perception, reasoning, anger, etc., to a mere body; yet it is almost equally difficult to ascribe them to what is left when the bodily factor is ignored." This joint hypothesis, he believes, would overcome most of the difficulties encountered by both the materialist and the interactionist theories in the forms hitherto propounded. Further, he thinks that there are many phenomena now apparently established by 'psychical research' —notably the phenomena of mediumship and post-mortem apparitions—which point to some such conception.[1]

Discussing the theory in a later publication, he points out that it might be developed along one or other of three possible lines. (i) We might assume that there is only *one* psychic factor for all minds. This version, he observes, would in many ways resemble T. H. Green's theory of the one Eternal Consciousness and the many animal organisms. But, so far as I can see, this form of theory differs but little from the doctrine of psychic monism proposed by Bergson and James. (ii) Secondly (and this seems the form that Broad himself prefers) we might assume that there is at any one time a different psychic factor for every mind, and that the factor can exist only in combination with a body, though it might pass in immediate succession from one body to another. This, as he points out, in some ways resembles the situation in chemistry where certain radicals, like NH_4 and CH_3, have at most only a transitory existence independent of a compound. Such a version "has advantages for those who favour the theory of metempsychosis" (McTaggart, for example, and many Hindu metaphysicians)—a theory which Broad maintains should be taken far more seriously by Western philosophers.[2] (iii) Thirdly, he thinks we might perhaps consider a theory which combines features from both the foregoing suggestions, and assume that highly individual minds like our own might require each its own psychic factor, but that lower organisms, among which the individual differences are presumably trivial, share a single psychic factor: for instance "it might be held that there is but one factor for the whole species of earwigs".

In all three versions the basic assumption, it will be noted, is the same, namely, that "mentality is an emergent characteristic of a compound composed of a living brain and something else which is capable of persisting after death and of then entering into temporary combination with the brain of a 'medium'. . . . Such a theory is compatible with all the facts that are generally admitted; and it has nothing against it except a superstitious objection to dualism". I should certainly prefer Broad's theory in its second form to that of James and Bergson. Otherwise I see little or no advantage in the more distinctive modifications which he suggests. Indeed they seem open to the following rather obvious criticisms.

---

[1] *Op. Cit.*, pp. 535f., 651f.

[2] C. D. Broad, *Examination of McTaggart's Philosophy*, 1933, II, p. 639.

(i) To begin with, if we may judge from the few characteristics of the 'psychic factor' that Broad himself enumerates, it would appear that the factor can lay hardly any claim to the epithet 'psychic'. It is neither rational nor sentient: it cannot therefore of itself be conscious or aware of anything.[1] Broad himself even admits that, though it "produces no ordinary physical and chemical effects, there is no reason why it should not be *material*". Possibly this is why in his later book (*loc. cit.*, p. 604) he refers to it as 'psychogenic' rather than 'psychic'. But, so far as his own account goes, it would be just as reasonable to apply this label to the organism itself: actually it would appear to be the two together that are 'psychogenic'.

(ii) Apart from these negative characteristics we are told practically nothing about the intrinsic nature of the 'factor'. Thus, as another critic[2] has urged, it remains "a mere X", and is no more than "a reified description of the facts which it is postulated to explain". It supplies no basis for predicting further observable facts which might serve to verify it, and appears to possess no essential qualities of its own.

(iii) Professor Broad is almost equally vague about the psychic compound which the two components jointly generate—namely, "mentality or what we know as the 'mind'". If, as he seems to imply, the compound is postulated primarily to perform the functions assigned by other philosophers to the 'pure Ego', i.e., to act as the subject-term to the various cognitive relations, then it is not easy to see why a 'Mixed' or 'Compound Factor' should be able to achieve this more effectively than a 'Simple Factor'.

Parallelism and Panpsychism. In the earlier discussions of the mind-and-body problem attempts were made to avoid the difficulties of materialistic theories on the one hand and animistic theories on the other by adopting a third more or less non-committal hypothesis. "The formula used for this purpose," says Stout[3] (its most influential champion), "is that of psychophysical parallelism." To every conscious state there corresponds a brain-state, and vice versa. "The chain of psychical causation runs parallel with the chain of material causation, but the two do not interact." If any further explanation of this 'pre-established harmony' is required, we must, he says, fall back not on a scientific but on a metaphysical hypothesis, namely, one which insists on an "idealistic view of all material phenomena". To the parallelistic formula as thus stated the most obvious answer is that not only are many brain-states unaccompanied by any corresponding conscious state, but also, as we have seen, there are more numerous conscious processes which do not have, and apparently could not have, any analogue in the chain of physical processes. For the rest, nearly all the arguments which weigh against epiphenomenalism weigh equally against this form of psychoneural parallelism.

The doctrine of parallelism, however, can be put in a wider and in some ways more attractive shape. Of the many versions of this broader theory probably the most plausible is that of C. A. Strong. In a volume with the title *Why the Mind has a Body?* (1903)

---

[1] Broad at times seems a little hesitant on this point: "it remains possible", he says, "that it *may* have some of the lower factors of mentality, such as sentience; but I do not see anything in the facts to suggest this hypothesis" (*op. cit.*, p. 651).

[2] Professor C. J. Ducasse *ap.* F. A. Schilpp, *The Philosophy of C. D. Broad*, 1959, p. 403.

[3] *Manual of Psychology*, 1904, pp. 50–57. In the latest edition of this work (1938) revised in collaboration with the author by Professor Mace, strong arguments are urged against the doctrine of parallelism as previously formulated; and, in a new section, the doctrine of interaction is treated far more sympathetically.

Strong rejects the ' instrumentalist's reply ' as formulated by James and Bergson, and puts forward a doctrine of panpsychism which he has since worked out in considerable detail.[1]

Strong's ultimate aim is to provide a theory which will be compatible with a modified metaphysical doctrine, which he calls ' critical realism '.[2] The older school of ' new realists ', who had risen in revolt against the ancient idealistic creeds, maintained that the sense-data, which are ' given ' in ordinary sense-experience and are commonly regarded as ' contents of consciousness ', are in fact components of the physical world: grass is really green, tomatoes are really red. To this Strong rightly replies that, if ' givenness ' is an intrinsic character of the physical objects we perceive in sensation, it would be difficult to account for the 'non-givenness' of those physical objects which nobody perceives. To call in a *deus ex machina* like Berkeley's ' omniscient all-perceiving deity ' would be no genuine solution. ' Givenness ' therefore must be the essential characteristic of a non-physical ' vehicle ' which he terms a ' psychic state '. The ' self ' is made up of such ' psychic states ', and must be considered " as a substantive rather than a separate substance ". The self or ' psyche ' thus consists of ' mind stuff ', and " mind stuff exists in space as well as in time ". Moreover it has no unity beyond the unity of the brain. The apparent paradoxes involved in this somewhat complicated theory are overcome by invoking the ancient ' double-aspect ' doctrine: all reality has both a physical and a mental aspect. Thus " the body is the appearance of the psyche to the outer senses "; and the psyche itself is " at once psychic and extended ".

Except in its rejection of instrumentalism in the narrower sense this theory closely resembles that of James. Like James, Strong believes in an all-embracing sea of mind stuff, which is apparently subdivided into separate individual ' psyches ' solely because of its association with separate individual brains. Evidently most of the criticisms we have against James also hold against Strong. To examine the merits of the various metaphysical doctrines used to eke out the rival psychological hypotheses—idealism, realism, pragmatism or the like—falls outside our present scope.

IV. SUMMARY AND CONCLUSIONS

A Suggested Model. By way of summing up the results of the foregoing arguments let me now attempt to put together my own conception of a ' mind '.

(1) First of all, I regard it as a ' continuant ' in Johnson's sense: it is not *merely* an ' existent ' or set of existents (for that would allow us to treat it as no more than a complex event or system of events); it is a particular type of existent, characterized by a certain degree of permanence. And in that rather restricted sense I should be prepared to describe the mind as a ' substance '. This first qualification rules out the conclusion which James reaches at the end of his long discussion about the ' stream of consciousness ' and the ' nature of the self ', namely, that " if the *passing thought* be the directly verifiable existent, then that thought is itself the thinker, and psychology need not look beyond " (*Principles*, I, p. 401).

(2) Secondly, the mind possesses a certain degree of *unity and independence*. It must in my view be a unity, because it is to serve as the basis of the individual's personal consciousness, and that, as we have seen, has unity as one of its most distinctive characteristics. It is independent both of other minds (i.e. of the bases of other persons' consciousness) and of the physical organism with which it is associated. I fully agree that the evidence provided by ' cross-correspondences ', and still more by reports of alleged apparitions after death, is

---

[1] In *The Origin of Consciousness*, in the later *Essays on Natural Origin of the Mind*, and finally in an important article in *Mind* (XL, 1931, pp. 217f.). It should be added that Fechner, as every student is aware, advanced a doctrine of psychophysical parallelism which included both the forms discussed and endeavoured to support it by reference to his famous law relating psychical sensation to physical stimulus.

[2] *Essays in Critical Realism*, 1920, pp. 225f.

far from convincing. Nevertheless, I hold with Broad that there is perhaps just sufficient to make it slightly more probable than not that something ‘psychic’ may survive after death, and continue to exist independently of any material body or brain. To many no doubt the issue will seem a matter for faith rather than for scientific proof. In any case it is important enough to forbid us at the very outset staking everything on a theory which would completely rule out any such possibility.

(3) Thirdly, the mind, like other ‘substances’, has certain *dispositional properties*. Of these, the most distinctive—the *differentia*, as the older logicians would have said—is the capacity for consciousness in the narrower sense of the term, i.e., for cognitive ‘awareness’. It is this third qualification that makes the ‘factor’ or ‘entity’ which we are postulating specifically ‘psychic’. ‘Awareness,’ as I have argued elsewhere, is a simple and unanalysable dyadic ‘relation’: as such it requires not only an ‘object’ (i.e. a sensum or an image) but also a ‘subject’. Or, to adopt Broad’s way of putting things, it involves an ‘act’ of prehension; and an act of prehension implies an agent which does the prehending. The ‘mental event’ is simply this ‘act’: if mental events did not happen to mental substances (as physical events happen to physical substances), then, as he observes, no event could consist in the prehension of anything.

But I hold that there are other types of cognitive relation besides ‘awareness’: cognition in my view comprises not only (i) ‘awareness’ or direct apprehension of (*a*) sensa and (*b*) images; it also includes (ii) intuitive forms of indirect (or referential) cognition, as in (*a*) perception and (*b*) memory, and (iii) discursive forms of indirect cognition, as in (*a*) judgement and (*b*) reasoning.[1] None of these forms of cognitive relation—the higher forms, as they are sometimes termed—is reducible solely to one of the lower or simpler relations. Together they guide or steer the activities of the embodied mind; they are, in current jargon, ‘cybernetic’. But I also hold that consciousness in the broad sense includes other relations besides those that are strictly cognitive. It includes certain affective or evaluative relations (being pleased with an object, etc.). Hence among the various dispositional properties of the mind I should place a capacity to appreciate values (and this in my view forms one of the most important objective characteristics of the universe we live in).[2] This property enables us to accept the everyday assumption of personal responsibility, and provides another striking peculiarity which distinguishes a mind from a mere material substance. Finally I hold that consciousness includes the possibility of certain conative relations. These last two sets of relations form the basis for the familiar theories about orectic or hormic ‘energy’. Thus the mind as I conceive it includes not only a system of *directive* dispositions but also a system of *dynamic* dispositions.[3] The nature of these ‘dispositional properties’, and of others equally distinctive but more specific, will be expressed in terms of what are sometimes called ‘scientific laws’, i.e., statements about general tendencies of the form “*if* so-and-so, *then* so-and-so”. These laws will plainly be quite different from the laws describing the dispositional properties of material substances, and could not possibly be derived or deduced from them. It is the business of psychology to discover and formulate such laws.

---

[1] The evidence for this is not merely introspective but statistical: see Burt, ‘The Structure of the Mind: A Review of the Results of Factor Analysis,’ *Brit. J. educ. Psychol.*, XIX, 1949, pp. 101–111, 176–199.

[2] Mr. Gregson and others (see this *Journal*, XIV, pp. 72–75 and refs.), have challenged this assumption on the ground that ‘value’ is not an objective quality and certainly not a concept that is permissible in a natural science. In reply I will only say that I accept as completely convincing the arguments brought forward by Professor A. E. Taylor to show that “the supposed absolute disjunction of fact and value is merely a mistaken prejudice” (see his Gifford Lectures on *The Faith of a Moralist*, pp. 46f.). Without such an assumption a psychology of delinquency and crime would become impossible.

[3] See Burt, ‘The Factorial Study of the Emotions’, ap. M. L. Reymert, *Feelings and Emotions*, 1950, pp. 531–551.

(4) It is evident from this brief review of its various properties that a mind, so far from being a '*simple* substance' must have a distinctive *structure*. Further, I believe that each individual mind will not only exhibit the general structure common to all minds, but will also develop structural peculiarities of its own. This would imply that the mind possesses a disposition to acquire dispositions, i.e. to form what would ordinarily be called memories and character-traits.[1] Could this conclusion be empirically established, we should then avoid two incidental difficulties: first, that a mind endowed solely with general properties would, if it survived, be wholly impersonal; secondly, that, if personal minds are sufficiently permanent to survive after death, they must for the same reason have existed before birth—a notion which seems to me far less plausible than it does to Broad and McTaggart. Indeed, since scientists are nowadays prepared to contemplate as a tenable hypothesis the continual creation of atoms (as the Hoyle and Bondi 'steady state' cosmology requires), there can be no reason why we should not also imagine—still merely as a hypothesis—the continual creation of minds. At some stage before birth we should then have to suppose that an individual mind emerged, as it were, first in the form of what Heisenberg has called a 'potentiality', i.e., as a 'mere seed, germ, or bud'; and personality would be something that gradually unfolds and develops during terrestrial life.

There are at least three ways in which the foregoing conception differs from the Cartesian theory, as Dr. Smythies and Mr. Hudson represent it. According to that theory, "the mind or self is not only a simple substance, it is also an unextended substance, and is something of which in introspection I am directly aware". None of these features forms part of my definition.

(i) To begin with, the mind as I have described it is something complex, and it has a history. If so, I can no more be 'directly aware' of my 'mind or self' than I can of any physical object: the most I could directly apprehend would be the thinker or perceiver of the passing moment. Many psychologists indeed would hold with Hume that they have no knowledge whatever of any such thing as a self. After all, awareness is from its very nature an asymmetrical relation; and should we not agree with the Hindu philosopher: "The eye sees not itself, the ear hears not itself, and the knower knows not himself". In reply I should argue that it is the mind that sees and hears, not the sense organ; and, just as I can legitimately talk of loving myself or being pleased with myself, so, in a certain sense, I may legitimately talk of knowing myself. There is no reason why an asymmetrical relation should not also be reflexive. Yet even so, I readily admit, my knowledge of myself must be more like my perceptual knowledge of a material thing—the clock upon my mantelpiece, for instance—than a direct or intuitive awareness of a permanent ego.[2]

---

[1] See the experimental work reported by McDougall and May Smith, *Brit. J. Psychol.*, X 1920, pp. 143f.; also *Body and Mind*, pp. 334f. Here, however, I must confess, the evidence so far available appears to my mind decidedly weak: it is quite conceivable that the special structural characteristics acquired by the individual may be merely acquired structural modifications of the individual's brain.

[2] Descartes' own account certainly seems quite unconvincing. "Of every substance," he says, "there is one principal property which constitutes its nature or essence. Thus extension in length, breadth, and depth constitutes the nature of corporal substance, and thought the nature of thinking substance" *Principles of Philosophy*, sect. LIII). In sect. XXXII he says: "there are only two modes of thinking, viz., the perception of the understanding", which includes "to perceive by senses" as well as "to imagine and to conceive" and the action of the will. However, in the *Meditations* perception and thought are contrasted. Perception, we are told, is commonly thought to be "an attribute of the soul" like nutrition and locomotion (the reference is to the Aristotelian conception of the *psyche* as adopted by the scholastics; nevertheless, "perception, like walking or taking nourishment, is impossible without a body" but thinking does not need a body. "Here therefore I discover what properly belongs to myself: I am, properly and precisely speaking, only a thinking thing, that is a mind (*mens sive animus*)." Animals do not think; therefore they have no minds.

(ii) The Cartesian assumption that the mind is a substance with neither location nor extension in space apparently forms the chief reason why so many psychologists reject the concept of mind in favour of a conscious brain which possesses both location and extension. With Russell and his disciples they maintain that my consciousness is ' in my head '. This seems rather like debating whether or not gravitation is ' in the earth '. I suspect that the psychologists in question picture the brain as a *tangible* thing in *perceptual* space. But the physical objects of science are not in perceptual space but in physical space or rather space-time; and for the physicist space or space-time is not an observable container in which things are visibly extended, but merely a name for a particular system of relations. Hence, much as I dislike using an ambiguous preposition like ' in ' of mind or consciousness, I see no reason why the mind should not be ' in space ' (in the physicists' sense) or ' extended ' (as I suppose a field of force is ' extended '). But the issue seems largely verbal and little more than a gratuitous pseudo-problem.

Interactions. The dispositional properties of the mind, like those of other substances, will be displayed by its interactions. The interactions of purely physical substances may be described as ' physico-physical '; and include both inter-physical processes (such as the interaction of an acid with a metal, or a proton with an electron) and intra-physical processes (as in beta-decay, or in growth and senescence, though most intra-physical reactions could probably be reduced to inter-physical reactions between the constituents of various compounds). Nearly all physiologists postulate interactions that may be described as physico-psychical, like those involved in sense-perception; but they apparently hesitate to accept the notion of psycho-physical or psycho-kinetic interactions, such as most laymen believe to occur in the case of volition. Both these types, however, are essential to my theory. In addition I assume at least the *possibility* of inter-psychical reactions, as in telepathy, and of intra-psychical reactions, as in reasoning (when it forms the ground for choice) and possibly in memory (as in the mode of recall postulated by Russell's theory of ' mnemic causation '). A strong confirmation of my general hypothesis would therefore be provided if a series of experimental or statistical investigations could be planned such as would yield positive evidence in favour of these additional types of interaction.

The Mind as a Field. To psychologists brought up, as so many of us still were, on the elementary notions of classical physics, the idea of an immaterial something, endowed with dispositional properties having no physical basis for their support, must sound almost as incredible as the idea of an immaterial something, endowed with electrical properties, would have seemed to the physicist at the outset of the nineteenth century. Electricity they thought of as a material fluid flowing through a physical wire. Maxwell pictured Faraday's electrical and magnetic field as a system of stresses and strains in the " contiguous particles of a universal medium ". It is only since Einstein " finally abolished the aether " that physicists have grown accustomed to the notion of a ' field ' without supposing that there must be some underlying physical medium to explain it.[1]

---

[1] In its simplest form the physicist's notion of a ' field ' may be explained by considering the case of the gravitational field. At any given point the attraction which results from a given distribution of mass can be expressed in terms either of a force which would act on a unit point mass placed there, or of a gravitational potential existing at that point and satisfying certain differential equations. The potential is literally ' potential ' in the sense that it does not describe a material property of the field, but may have a value in empty space: it thus describes a *dispositional property*, namely, the force that *would* be exerted *if* a unit point mass were introduced into the field. In this sense, as Hertz observes, " the field-strength is a reality which exists even when the reacting bodies are removed ". Planck uses almost exactly the same phrase: (*Theory of Electricity and Magnetism*, 1932, p. 110); and Reichenbach applies the word ' substance ' to describe the field (*Atom and Cosmos*, 1932, pp. 124, 187f.). Note that the force has a *direction* as well as a *strength*; (every schoolboy is familiar with the experiment in which, say, the north pole of a magnetized needle transfixed vertically through a floating cork, moves in a definite direction along a curved path away from the north pole of a bar magnet around towards the south pole).

Now if we think of the dispositional properties of the mind as psychic ' tendencies ', analogous to the physicist's ' potentials ' or ' forces ', we can regard the mind as a kind of ' field ' existing in the neighbourhood of an individual brain, much as the physicist conceives of an electromagnetic field existing in the neighbourhood of a set of electric currents. The tendencies would exhibit both dynamic and directive characteristics. Such a field would of course differ from the more familiar types, just as electromagnetic fields differ from gravitational, topological from metricized, and quantized from unquantized. But the scientist of today shows no hesitation over postulating new types of field when they seem needed to interpret phenomena which would otherwise escape unexplained.[1]

The Double-Filter Theory. A plausible hypothesis, not unlike the so-called ' instrumental theory ' of James and Bergson, might, I think, be developed, if we consider the biological implications of a classification of cosmic processes which has often been involved in discussing the newer physical theories. We can divide the whole scale of spatial and temporal quantities with which the different sciences deal into three broad orders of magnitude. We can then distinguish between what I will call macrophysical, mesophysical, and microphysical phenomena respectively.[2] By mesophysical phenomena I mean those in which the spatial dimensions, the time-intervals, and the velocities do not greatly exceed or fall short of those we encounter in our everyday perceptual experience. They are the familiar phenomena which are regularly described in the language of ordinary life and in terms of naïve realism. Here too belong not only the problems, the concepts, and the models of classical physics, but also (be it noted) the observations made in every attempt to verify a physical hypothesis in the laboratory or elsewhere, no matter which class of phenomena is under investigation. By microphysical phenomena I mean processes below this range, particularly those of the atom, the atomic nucleus, and the various elementary particles. At this level entirely new concepts and models are necessary—those associated with quantum theory. By macrophysical phenomena I mean phenomena in which the spatial dimensions and the time-intervals are literally of astronomical dimensions, and the velocities so high as to approach the speed of light. Here belong the novel concepts and problems of cosmology, and more particularly those dealt with in the theory of relativity.

In modern investigations of microphysical and macrophysical phenomena it has proved extremely difficult to ignore the interactions of the conscious observer with the system under investigation. Consequently the borderline between physics and metaphysics tends to disappear,[3] and both Occam's rule of parsimony and the postulate of simplicity are apt to break down. Contrary to all expectation the number of basic 'entities' has steadily increased; and the equations have become more and more complex. The simpler formulae of classical physics turn out to be only working approximations obtained in those intermediate cases where the mesophysical phenomena are sufficiently large for Planck's constant to be neglected and the velocities are sufficiently small for the propagation of light to be treated as instantaneous. Thus "the rigorous laws of physical science" to which Watson, Hansel, and others would like psychology to conform are no more than approximate probabilistic laws.

---

[1] Not only telepathy (if it exists) but also every form of cognition is essentially a mode of action at a distance. Now in physics a common reason for postulating a ' field ' is to furnish a convenient model for interpreting action at a distance: and in my view this is one of the many advantages of adopting a similar model in interpreting the phenomena of consciousness.

[2] The first to emphasize and label the basic distinction seems to have been Eddington: he distinguishes what he calls ' macroscopic systems ' which " broadly speaking are covered by relativity theory " and " microscopic systems covered by quantum theory ". The term, however, is not altogether appropriate, since the 'entities' which he specifies—protons, electrons, photons, etc.—are submicroscopic. The term ' macroscopic ' includes what I have called ' mesophysical '.

[3] Von Neumann's brilliant examination of the process of quantum mechanical measurement plainly demonstrates that " the laws of quantum mechanics, with its various implications, cannot be formulated without recourse to the concept of consciousness " (J. von Neumann, *Mathematische Grundlagen der Quantenmechanik*, Engl. Tr., 1955). Heisenberg puts the situation very clearly. " The mechanics of Newton, and of classical physics constructed after its model, started from the assumption that one can describe the world without speaking about ourselves. This seemed almost a necessary condition for natural science. But the situation changed as a result of quantum theory. . . . The system which is treated by the methods of quantum mechanics remains part of a much bigger system: it could not in fact belong to the world of observable phenomena if it had no interactions with this bigger system of which the observer is a part." Thus, contrary to the assumption of the Newtonian physicist, "natural science does not simply describe and explain nature; *it is*

Now the terrestrial world in which the human mind as we know it has presumably evolved is essentially a world of mesophysical dimensions. In order to survive in such a world the size of man's body cannot exceed or fall short of certain limits, and the range and speed of his movements are correspondingly restricted. A sensory awareness of phenomena greatly beyond the limits thus imposed would be a hindrance not a help. As a result the brain and sense-organs with which natural selection has equipped him are highly selective. The potential stimuli arising in his terrestrial environment are thereby filtered and sifted before they are allowed to enter his mind. Radiations exist well beyond both ends of the visible spectrum: the 'waves' used for wireless transmission are 500 yards or more in length; X-rays are only about 1/10,000 inch in length; and gamma rays and cosmic rays less than 1/200,000,000 inch. If the eye could see the whole range of these invisible colours, the spectacle presented by the universe would be far richer and far more informative than it actually is. But with X-rays objects would seem transparent. Rays of a wavelength much longer than light (heat rays, for example) turn corners, and would be of little use for direction-finding. For these and other reasons human vision is therefore limited to a narrow band between 16/100,000 and 32/100,000 inch. Man's chemical senses are poorer still; and he has no sense-organs that enable him to recognize electrical or magnetic fields. Nature allows him to perceive only those properties of the material world which are really important for his survival as a land animal—sustaining and defending himself and his family on a very restricted portion of a very unusual planet; a glance at the vast number of instruments carried on an aeroplane and the innumerable precautions taken to protect an astronaut, is sufficient to demonstrate how scantily the human body is equipped by nature for appreciating conditions other than those for which it was actually evolved.

Man's interpretation of the information which his senses supply has also to be limited to the barest essentials. For quick and efficient action the universe must appear extremely simple, and man must think of it in simple ways. Day-to-day life must seem so plain and straightforward that "the wayfaring man, though fool, cannot err therein". Much of his perceptual interpretation his brain automatically supplies of itself. And, were he forced to recognize how incomplete and subjective his picture of life really is, he would hesitate over every action: a race of reflective Hamlets would become extinct with the very first generation. Consciousness itself therefore has to be as transparent as a glass window, so that the man of common sense, like the behaviourist, can ignore its very existence.

The brain, then, in transmitting information about the cosmos to the conscious mind, must permit only what is absolutely indispensable to percolate through. But equally, I believe—though here we are on still more speculative ground—the brain must also impose restrictions on what the conscious mind itself is capable of performing. Faced with the problems of survival or of advancement for himself and his family, the practical man,

---

itself an interplay between nature and ourselves" (W. Heisenberg, *Physics and Philosophy: the Revolution in Modern Science*, 1959, pp. 74–5; cf. pp. 153–4). As he sums it up elsewhere, "the conception of objective reality has evaporated into the transparent clarity of a mathematics that represents no longer the actual behaviour of elementary particles, but rather our *knowledge* of this behaviour" (*Daedalus*, LXXXVII, 1958, p. 100).

primitive or civilized, seems compelled to think first—and often solely—of the significance of things in that very limited context: they are viewed as means, never as ends in themselves.

The poet and the artist, on the other hand, sees or claims to see them as possessing a significance of their own. For them " the meanest flower that blows . . . meadow, grove and stream, the earth and every common object ", have each its own peculiar value, and appear, not merely 'apparelled in celestial light', but glowing (so to speak) with an intrinsic meaning and a glamour of their own. At the same time, however, " it is his absorption in what he takes to be these deeper meanings that makes the artist seem at times abstracted and unpractical—an absent-minded dreamer lost in ethereal visions ".[1] Similarly the abstract scientist and the philosopher, who pursue truth for its own sake rather than for its practical benefits, have often appeared figures of fun to the man of common sense: Thales, gazing at the stars, falls into the well. But the supreme example of this contrast between the search for supposed intrinsic value and the pursuit of material satisfactions is the mystic. On a strictly materialistic view, such as Dr. Bolton advocates, we are bound to regard all mystical experiences as " the illusory effects of a semi-pathological condition ", and hence to reject their implications. And yet, when we consider how much modern civilization, and even science itself, owes to such people—to Plato and Plotinus, to Buddha and the anonymous Deutero-Isaiah, to Paul of Tarsus and to Augustine of Hippo—is it not a little rash to be so confident of the present stage of human knowledge that we can wholly dismiss what they claim to have discovered from their wider range of consciousness? Without necessarily committing himself, the psychologist must at least leave room for the *possibility* that their claims may be in some measure justified. Indeed, as William James maintained in his studies of such states, " the conclusion seems often forced on us . . . that our normal, waking, rational consciousness is only one rather limited type: round about it, parted by the flimsiest of screens, there lie potential forms of consciousness entirely different. We may go through life without suspecting their existence; yet no account of the universe can be final which leaves them disregarded ".[2]

Ordinarily, in the interests of survival, the brain shuts them out; but in certain more or less abnormal states[3]—states of dissociation, and the like—these latent capacities seem at times to force their way out and become highly active. Such states, however, are exceptional. The material and utilitarian outlook of civilized life almost inevitably forbids any diversion of attention to the remoter planes of cosmic existence. And for these reasons, we may imagine, as Dr. Tyrell has observed, " clairvoyance, telepathy, and other paranormal phenomena, like those of religious experience and mysticism, crop up far more frequently among primitive tribes and individuals, and seem all but extinct among the highly civilized, the highly commercialized, and the scientifically trained ".

Conclusion. Briefly then the hypothesis I suggest may be summed up as follows. The mind is a unique type of 'continuant', characterized by a certain unity and independence, and endowed with certain dispositional properties and a definite structure, which are essentially psychic. It thus provides a relatively permanent basis for consciousness. It is not to be interpreted as a resuscitation of the 'simple, unextended, indestructible substance' of Cartesian metaphysics. It resembles, not the traditional *anima* or soul, but rather what the physicist would call a 'field'. In the end research may possibly show that the 'continuant' has no greater permanence than the individual's terrestrial life—perhaps that it could not even come into existence apart from a special kind of physico-chemical sub-structure—a living organism or brain. At the moment, however, the wavering balance of evidence seems to point rather in the opposite direction. But the main advantage of such a conception, as compared with the purely physiological interpretations favoured by materialists, physicalists, and behaviourists, springs from the fact that it does not from the very outset preclude investigations concerned with conscious processes as such or with the alleged paranormal phenomena studied in psychical research; and it furnishes us with a far more convenient framework for describing the conscious processes of everyday life, and consequently for the various branches of applied psychology. Finally, let me repeat, the whole conception is presented, not as a theory fully confirmed by *a posteriori* evidence, but merely as a tentative model, suggested by certain recurrent phenomena, implying certain testable corollaries, and urgently needing to be verified, refuted, or modified by *ad hoc* empirical research.

---

[1] I quote from my chapter on the 'Psychology of Art' in *How the Mind Works*, where I have discussed this view of 'value' at some length.

[2] W. James, *Varieties of Religious Experience*, 1904, p. 427.

[3] Nevertheless, an abnormal state is not a *sine qua non*. John Buchan, for example, records that, as a rule, he had these rare quasi-mystical experiences with their 'wonderful clearness of mind' and the 'sudden relevation as in poetry or music', when he was 'very fit in body' (*Memory-hold-the-door*, 1940, pp. 120–21).

HIGH ABILITY AND HIGH ACHIEVEMENT

To the Editor, British Journal of Statistical Psychology

SIR,—In his recent discussion of social mobility Professor Burt appears to assume that his assessments of 'intelligence' are measurements of an intellectual capacity that is purely, or at any rate mainly, innate. I am aware that in previous articles he has already carried out an elaborate analysis of variance in the endeavour to show that, in theory, intelligence is essentially an inborn characteristic. But, in discussing such phenomena as social mobility, we are concerned not with psychological theory, but with the practical performance of actual individuals, and of individuals of particular types—the mentally subnormal and the mentally supernormal, whether or not (as we are asked to assume) they inherit their subnormality or supernormality from their subnormal or supernormal parents.

The educational system Professor Burt would have us establish implies a society in which class distinctions, instead of being reduced or superseded, are to be sharply accentuated. He displays a surprising faith in the predictive value of examinations and tests, and appears to claim for the 'gifted groups' discovered by their means a complete freedom from almost every type of blemish. As to the outcome, were such selective methods to achieve success, two views are possible. According to his view, they would result in a community arranged in a hierarchy of *intellectual* grades; presumably with a glorified public school for the intellectual élite at the top—in short, the rise of meritocracy. According to the view accepted by most contemporary psychologists and sociologists in America as well as in this country, they would create or perpetuate a community arranged in *social* grades, based on income, privilege, or sheer intellectual snobbery.

Since, as we all agree, 'gifted' pupils are commonly the offspring of 'gifted' parents, it is after all only to be expected that we should find them chiefly in professional classes (or the equivalent ranks in commerce or industry). In these upper social classes the parents are obviously in a position to supply their offspring with far richer opportunities both at home and at school for developing 'gifts'. Giftedness is thus not a biological endowment but the superficial effect of the unfair advantages that accrue to members of the privileged and well-to-do classes. Galton's own collection of pedigrees and family histories are far from proving his hereditarian thesis. All they really demonstrate is that the special achievements of men of genius are predominantly the result of the special facilities and the special stimulation afforded by their environment. I suppose the most obvious hall-mark of intellectual achievement is a Fellowship of the Royal Society. Yet consider how many of them—from the first President, Lord Brouncker, and the Honourable Robert Boyle, 'father of chemistry and uncle to the Earl of Cork' down to quite recent times—have been noblemen, or members of wealthy and aristocratic families, or close friends of the Court like Pepys and Evelyn. Surely the work of Dr. Fleming, Dr. Heim, Dr. Floud and others has abundantly shown that the role formerly attributed to heredity by Pearson, Terman, and other members of the Galton school was greatly exaggerated, and that individual differences in ability or aptitude are due far more to social and environmental influences than to genetic conditions.

A. L. HARRISON

SIR,—In a recent article (this *Journal*, XIV, 1961, p. 13) Professor Burt cites figures showing that nearly half the children of high intelligence fail to achieve an occupational status equivalent to their capacity; and in other papers addressed to educationists and teachers (*Brit. J. educ. Stud.* VII, pp. 114f., *Educational Research* II, pp. 241f.) he complains that, in our eagerness to do all we can for the mentally handicapped and the large mass of average children, we have neglected to devote sufficient attention to those who are of outstanding ability. May I suggest that for this psychologists themselves are largely to blame? Many of them seem to imagine that the importance of environment, and of environmental improvements, can only be brought home if the importance of heredity and genetic conditions generally is belittled or denied. And those who believe in the importance of both fail to present their evidence in a way which will convince the practical worker and the general public. The highly complicated algebraic proofs offered by Fisher, Burt, Howard and others are quite incomprehensible to the average teacher or educationist, and probably I suspect to most psychologists. Would it not be possible to give a few straightforward arguments, based on plain and concrete evidence, without all these mathematical complications?

And why do psychologists persist in treating the problems of the supernormal as if they were of the same general type as the problems of the subnormal? Throughout their discussions both issues seem to be regarded as no more than special aspects of the broader and vaguer problems of 'general intelligence'. Even when they turn from their curves and their calculations to talk about actual individuals, we still hear far more about the subnormal than about the supernormal. We have books on 'the backward child' and books on the 'mentally deficient', but no book on the 'gifted child' as such. It is principally for these reasons that the question has been so widely ignored by the public at large. Meanwhile, the critics of the theories Professor Burt so strongly advocates—the 'general factor', the genetic contribution, and wide range of innate differences—formulate their objections in everyday, popular language which all of us can understand, blandly ignoring the scientific evidence based on statistical analyses and silently passing over the statistical evidence and the vast array of quantitative data, no matter how thoroughly it has been tabulated, analysed, and reduced to symbolic equations. As a result their views are widely accepted, not because their arguments are conclusive, but merely because they alone are intelligible.

I suggest that, in the present state of psychology, the function of a statistical journal is not merely to appeal to the statistical specialist, but also (if I may use the word) to *popularize* statistical methods and statistical conclusions by recasting them in non-statistical form.

J. F. SOMERVILLE

Vol. XIV Part 2 The British Journal of Statistical Psychology November 1961

CRITICAL NOTICE

THE PSYCHOLOGY OF PERCEPTION

The Problems of Perception. By R. J. HIRST. London: George Allen & Unwin; New York: The Macmillan Company, 1959. Pp. 330. 30*s*. net.

A glance at the writings of present-day physicists is sufficient to show how the recent revolutions in physical theory have revived in a sharpened form all the time-honoured problems of perception, and revealed a host of new puzzles and paradoxes crying out for investigation. As Dr. Hirst rightly insists, the study of perception and of what he calls 'perceptual consciousness' is primarily a task for the psychologist; and yet, as the physicist's interest in the problem has grown, the interest of psychologists has faded and flagged. Fortunately, the current changes in philosophy have helped to keep the problem alive; and it is greatly to be hoped that Dr. Hirst's lucid and critical review may induce psychological investigators to turn their attention once again to a topic which formerly occupied the opening chapters of almost every textbook.

It is true that in the thirties the conflicting theories of post-Watsonian behaviourists stimulated a number of fruitful researches on some of the processes underlying perception but they commonly treated it, not as a unique type of cognition, but as merely a mode of learning or of communication, of verbal report or neural functioning. Perceptual consciousness as such was assumed to fall outside the legitimate pale of scientific inquiry, for the simple reason that all forms of consciousness are purely private phenomena, while science can deal only with what is public.

Statistical psychology, however—at any rate in this country—has been less affected by these psychophobic trends; and I venture to think that the approach which the statistical investigator habitually adopts, and the outlook that he acquires from his somewhat peculiar methodology, may help him to reach a solution which in some ways may prove more satisfactory than any of the conflicting theories advanced hitherto by the philosopher, the physiologist, or the experimental psychologist. They tend to oversimplify. Their theories are framed in terms of mutually exclusive, clear-cut categories. They ask such questions as: are our percepts private or public, objective or subjective? are sensory qualities physical or mental? And their central problem is commonly the problem of causation: are our percepts generated by the physical stimuli, by the neural structure of the brain, or by something called an 'ego' or a 'mind', and do they have any causal efficacy of their own? The most important outcome of this mode of approach is 'the causal theory of perception', which, so Professor Price tells us, has become 'the official foundation of natural science'—of natural science, that is, as understood by physiologists, doctors, and the majority of psychologists.

The statistical psychologist thinks in terms of variations and covariations rather than of causal properties or verbal classification. He sees phenomena as part of a multivariate pattern; and he asks, not which is the one essential factor, but which of the many contributing factors are the more important. But the most vital difference, I fancy, is this. The traditional outlook of the philosopher, and of the school of classical physics in which most of the earlier behaviourists were trained, leads them to seek conclusions that are absolutely sure and certain. If the various forms of perceptual consciousness cannot be operationally defined, or if statements about it cannot be objectively verified, then such notions and such statements must be banished from psychology as a reputable branch of knowledge. Or again, if, as some would hold, sensory phenomena are the only things of which we can claim infallible knowledge, then, according to a rival school, sensory phenomena are the only things that we can validly assume to exist. Time after time, as we learn from Dr. Hirst, the phrase " such a view is tantamount to denying the possibility of knowledge " is treated as a cogent *reductio ad absurdum*. The statistical psychologist adopts a more modest standpoint. He distrusts all claims to certainty, and is perfectly content to assess probabilities. He presents his own solution, not as an authentic description of what really happens, but merely as a hypothetical model—usually as just one of several possible models, each appropriate for some particular purpose.

With much of this, I suspect, Dr. Hirst would readily agree. In his opening chapters he criticizes the epistemologist for " the attempt to say with certainty what the object of awareness is "; and he deprecates the familiar type of syllogistic argument that takes as its premises " this or that (e.g., a particular type of observable datum or sensible quality) is indubitably present " and then seeks to deduce equally " indubitable conclusions ". Nevertheless, in the body of his work Dr. Hirst himself seems to fall back very largely on armchair deduction, and the stock academic *Gedankenversuche* about speckled hens, elliptical pennies, sticks in the water, and the reflected duplicates that philosophical Alices seem so surprised to discover when they gaze into the looking glass. On the other hand, he makes comparatively little use of findings reached by planned observations or recent experimental researches.

At the outset of his book Dr. Hirst largely follows in the steps of his predecessor, Professor Price.[1] Like him, he begins by considering the difficulties involved in " the usual statement of ' naïve realism ', the alleged creed of plain men ", which (as he very truly remarks) is " not at all fair to the plain man ". He then devotes two closely reasoned chapters to a drastic, but to my mind not wholly convincing, criticism of the doctrine of ' sense data ', as formulated by Moore, Broad, Ayer, and Price himself. From this he turns to the more important contemporary theories—first, that of ' phenomenalism ' as formulated in the later work of Russell and the earlier work of Ayer, secondly, that put forward by the champions of the ' linguistic approach ', notably Ryle, and thirdly, the modern variants of the ' representative theory ', more particularly those developed by neurologists, like Eccles, Russell Brain and J. R. Smythies. Each of these interpretations, he concludes, is open to grave objections, springing chiefly from untenable assumptions about the relations between consciousness and brain.

This leads him to an interesting digression on the problem of ' mind and body '. The interactionist theory is rather summarily dismissed. The behaviourist view as formulated in Ryle's *Concept of Mind*, together with his forthright rejection of introspection as a source of evidence, is subjected to a full and highly critical examination, with the result that Ryle's ' main thesis ' is rejected as ' quite incredible '. In place of it Dr. Hirst propounds what he calls an ' Identity Hypothesis '. " The person," he maintains, " is a unity, not an association of two distinct substances . . . An experience and a pattern of brain process are the same event differently revealed." Thus " the mental and the cerebral activities of the person are really two aspects of an identical activity ".

This version of the ' double-aspect theory ' is supplemented by a ' Sentience Hypothesis '. To justify it Dr. Hirst draws a sharp distinction between " percepta " (i.e., " how the object actually appears to the observer ") and " percipienda " (i.e., " those theoretically apparent properties which, from our knowledge of its nature and position, etc., the object should appear to have "). The discrepancies between the two he attributes to " modifications by the percipient of a basic awareness or ' sentience ' which corresponds to the theoretically apparent properties ". The nature of these various modifications is then investigated in a chapter on ' the psychology of perceptual consciousness '. This I find disappointing. The heading leads one to expect an up-to-date survey of psychological researches, old and new, on the various problems involved. Instead there are one or two incidental references to textbooks by Woodworth, Koffka, Hilgard, and Gibson,[1] a somewhat cursory discussion of ‘stereoscopic vision’ and ‘figure-ground and constancy’, and an ingenious version of the ‘television tube analogy’; and from these a theory of four basic types of modification is derived, viz., (i) the unconscious selection of certain parts of the percipienda through involuntary attention; (ii) a fusing of sentience with imaginative or conceptual elements; (iii) the unconscious use of cues (e.g., for visual depth and object constancy); (iv) various effects produced by learning and experience. Dr. Hirst (as he frankly owns) has found some difficulty in explaining how these various influences transform ‘sentience’ into ‘percepta’. He tries in turn such notions as ‘adjustment’, ‘organization’, ‘interpretation’, and the like, and thinks none of them wholly satisfactory. But the complex process he has in mind—which, as he says, usually includes the attribution of ‘meaning’ to the otherwise meaningless ‘sentience’—is what older British psychologists would have called ‘apperception’, and have already analysed in considerable detail.[2]

---

[1] H. H. Price, *Perception* (1932).

The concluding chapter opens with a ‘re-examination of the physiological evidence’. Then, in the light of this, Dr. Hirst states and defends his ‘final account of perception’. “Perception,” he decides, “is a relation between a person and a public object, in which a mode of active experience, ‘perceptual consciousness’, is caused in him by the stimulation of his sense-organs by the object or by emanations from it.” Both ‘perceiving’ and ‘perceptual consciousness’ present two aspects. The inner aspects of each consist in an awareness of an object or scene as external; but their outer aspects are different. The outer aspect of perceiving is a causal chain from object to brain *via* the sense-organs; the outer aspect of perceptual consciousness is “the same brain activity as in the perceiving, or closely similar to it”, but “it does not include any processes outside the body”.

This abbreviated account hardly does justice to the complexity of Dr. Hirst’s rather elaborate theory, or to the reasoned arguments by which he seeks to support it. But it will be sufficient to show that, while giving due credit to results obtainable by introspection, he is anxious not to depart too far from the general trend of psychological opinion since the behaviourist movement began. In the first place he seems unreservedly to accept the prevailing view that the essential characteristic of the objects of perception is that they are ‘interpersonal’ or ‘public’. Secondly, he holds that the critics of the traditional ‘bifurcation of nature’ have fully established their case, and is therefore anxious to keep to a hypothesis which will avoid the “faults of dualism”, whether epistemological (as in the ‘representative theory’) or psychophysical (as in the ‘interactionist theory’).

His attempt to break loose from epistemological dualism has not, in my view, been altogether successful. What he calls ‘percipienda’ differ from what other writers call ‘sense data’ only by the fact that Dr. Hirst describes them as ‘theoretically’ apparent. Thus, after all, his ‘final theory’ comes very close to that commonly offered by the more modern exponents of the ‘representative theory’. Indeed, as he himself observes, “I have to admit that the *content* of perceptual consciousness does at least *correspond* with the external *object*: I avoid the word ‘represent’ as too suggestive of pictorial reproduction” (my italics).

I fancy the difficulties that Dr. Hirst has so candidly acknowledged spring nearly all from the fact that he, in common with most of the contemporary philosophers whom he criticizes, still clings to a psychological account of ‘perceptual consciousness’ which is quite remote from the normal experience of everyday life. He and they take it for granted that the firmest starting-point for an interpretation of perception must be what he calls the ' physiological evidence ', or, as he puts it elsewhere, our knowledge of " the causal processes involved ". Light waves or sound waves emitted by the object eventually reach and stimulate the sense-organ; from the sense-organ electro-chemical impulses travel up the afferent nerves to the appropriate cortical areas. The immediate outcome (it is supposed) must therefore be a kind of ' sentience ' generated by the nerve-tissue, much as a latent picture is generated on the plate at the back of the photographer's camera. This latent sentience is then, as it were, developed and modified in the darkroom of the brain, so as to yield a perceptual ' content ' which ' corresponds to ' or ' symbolizes ' the external object.

---

[1] Readers who are interested in more recent surveys may be referred to D. W. Hammond’s *Psychology of Perception*, S. H. Bartley’s *Principles of Perception*, or, for a British study of the subject, D. E. Broadbent’s *Perception and Communication*.

[2] See more especially G. F. Stout, *Analytic Psychology*, II, pp. 110f. Stout, it may be noted, also uses the term ‘sentience’ (defined in his erudite fashion as ‘anoetic consciousness’) and adopts it to describe ‘sensory presentations’ before they have been modified by the processes of noetic synthesis’ (*ibid.*, I, p. 50).

Now I maintain that this is far from conveying an accurate impression of what really happens in ordinary everyday perception. In my waking experience I am aware of an ever-changing sensory continuum; and out of this complex and variegated background whatever I attend to at the moment stands forth as a visible, tangible, and more or less accessible object—' out there '. When I look at the bowl of fruit on the side table, I do not see a ' sentient ' reddish-yellow circular patch, ' modified ' into a fruity ' perceptum '; much less do I see a ' sense-datum ' or a sensory ' content ' that ' represents ' or ' corresponds to ' the fruit. What I see is a solid globular orange, just out of arm's reach. And I at once apprehend it intuitively as an independent and relatively permanent thing—an object of which I have, not mediate, but immediate knowledge.

No doubt, if we take as the basis of our reasoning what Dr. Hirst calls ' the physiological evidence ', we shall be tempted to conclude that " what is found at the end of so long and complicated a series of causal events, namely, the perceptual content, is unlikely to be the same as the remote object from which the whole transaction started "[1]. The argument would be perfectly valid if we could identify the ' perceptual content ' with the retinal image, or with an analogous facsimile " generated in the cortex " and then projected into the outer world. But, unfortunately for those who love simple causal explanations and shun anything that savours of action at a distance, what actually happens *is* an immediate and intuitive apprehension of the object as a concrete object seen at a distance from the very start, and not of a mere projected simulacrum. The colour, the shape, even the solidity are directly apprehended as essential characteristics of the thing itself; and we seem forced to recognize a unique kind of cognitive relation linking the percipient to the object perceived regardless of intervening space. And in point of fact the most that we are entitled to deduce from the physiology of sensation is that (paranormal cognition apart) the sensori-neural system acts as an indispensable instrument for perception.

The difference between the philosophical approach and that of the psychologist is clear. The philosopher is supposed to sit in his armchair, close his eyes, and deduce " from the known facts of physics and physiology " the character of the perceptual process. The unprejudiced psychologist opens his eyes, and describes—or should describe—what he actually sees, no matter how paradoxical the description may seem. If we could without more ado accept such an account at its face value, then " the red glow of sunset would be as much a part of nature as the molecules and electric waves by which men of science would explain it "[1]; and, I would add, the beauty and the truth of nature would themselves be as objective as the sounds and colours. With this interpretation the contents of the world around us turn out to be far richer and far more complex than the scientist for his own special purposes has commonly assumed. The physical objects of science are mere abstractions from the concrete objects that we perceive—bare skeletons, simplified geometrical or mathematical models, not the luxuriant, richly tinted, redolent, flesh-and-blood realities of our everyday life.

---

[1] This is the dualist's way of stating the conclusion (A. O. Lovejoy, *The Revolt against Dualism*, p. 49). Dr. Hirst contends that the mistake of the ordinary dualist is to " confuse the last stage of the process with the whole ". The dictum quoted in the text is a modernized version of a very ancient philosophical argument: " ' The effect ', it is said, ' must resemble its cause.' Hence, whenever what is transmitted is not a copy of the original emitter (as the scholastics insisted it must be in perception), then what is ultimately received will no longer resemble the emitter." This is an argument that is manifestly contradicted by the success of the telephone, of broadcasting, and of television. There what is ultimately received does resemble " the remote object from which the whole transaction started ". However, in perception—e.g. when I pick up and eat an apple from the bowl of fruit, what I see and feel and taste does not *resemble* the apple; it *is* the apple.

If I am asked, does this mean that ' objects ' still display their colour and their beauty when no real or hypothetical mind happens to be perceiving or thinking of them, and " when there's no one about in the quad ", then I answer that the question is for us quite meaningless, if not self-contradictory. All human knowledge is and must be relative. " Wir können die Dinge nur so denken wie wir sie vorfinden, nie so wie sie jemand vorfindet; wir können die Welt nur von dem Standpunkte aus denken auf dem wir wirklich stehen, nicht von einem Standpunkte aus auf dem wir überhaupt uns gar nicht stehend denken können, oder von gar keinem Standpunkte aus. Ich kann gar nichts absolut denken: im Stillen stelle ich mich noch immer als der Beobachter vor, der die Dinge sich gegenüber sieht."[2] The consequence of this inevitable reservation is to turn the doctrine I have been expounding from a kind of ' natural realism ' into a form of what has sometimes been termed ' objective relativism '.[3] By this I mean that, although what we know is necessarily relative to the knower, it is none the less objective—or at least more objective than subjective. To borrow the language of the statistical psychologist, the observer, as it were, serves as the ' origin ' for a particular ' frame of reference '; and what he perceives varies not only with the variations of the external object; it also varies with the intervening media, and the position or state of the observer's body and mind; but it is *most closely* correlated with variations in the object.

At this point, however, I think what has been called the ' physiological evidence ' ought to be supplemented by what may be termed the ' biological evidence '. Natural selection, in progressively evolving the instrumental apparatus of sense-organs and nervous system, has succeeded in ensuring, not that our knowledge should be perfect or complete, but merely that it shall be adequate for the practical purposes of mundane life under all normal conditions. In exceptional circumstances our ostensible percepts may become more or less confused or even definitely misleading. " Things," as the plain man, like the poet, quickly discovers, " are not always what they seem." In such cases the natural explanation—the one which the unsophisticated thinker usually adopts—is that the relation or process of cognition is itself a somewhat variable or flexible affair. Just as I can aim a dart or kick a ball more or less efficiently, so (it would ordinarily be said) I can discern objects more or less efficiently—the efficiency depending largely on the external circumstances and the condition of my own neural apparatus.[1] Dr. Hirst recognizes that in this respect common sense may have hit upon a tenable explanation; but, if I understand him rightly, he is doubtful how far such a theory, in its rather naïve form, can be rendered consistent with the physiological evidence, and therefore feels that it requires the supplementary hypotheses of 'sentience' and 'percipienda'.

---

[1] A. N. Whitehead, *The Concept of Nature*, p. 29.

[2] W. Petzoldt, *Das Weltproblem vom positivischen Standpunkte aus*, p. 142. (The pronouncement is really that of the physicist Mach, though the words are those of his philosophical disciple, Petzoldt.) As I have pointed out elsewhere, many if not most contemporary physicists would apply this qualification even to the objects of physical science. " Classical physics," says Heisenberg, " started from the belief—or should we say illusion?—that we could describe the world, or at any rate parts of it, without reference to ourselves." In modern physics we realize that " our descriptions cannot be completely objective ".

[3] The phrase is due to A. E. Murphy; but has usually been treated merely as a theory of perception, whereas I regard it as a statement about all human knowledge: it is also usually taken to imply that sense data are ' universal ', whereas every psychologist, I fancy, must insist that they are ' particulars '. Except for this last reservation what I have said brings my view a little nearer to that which was at one time advocated by Whitehead. But instead of his theory of the object as ' multiply located ' I should prefer to substitute a theory of the object as a pervasive ' field '. Faraday, it may be remembered, held that, according to his theory, " an electric charge is in a sense *everywhere* ". Some atomic theorists have used similar language about their ' elementary particles '. And generally, I believe, wherever traditional science and traditional psychology use the concept of ' substance ', the concept of ' field ' provides a far more useful model.

At this stage in the argument, however, I want to emphasize a distinction which hitherto has almost always been overlooked—a distinction that turns on the different ways in which our attention may be directed. Whenever our perceptions become confused or contradictory, our mode of attention alters; its focus is abruptly shifted from the object to the sensory characteristics. The sudden jerk produces a sort of mental dislocation. As a result the sense-datum is apt, as it were, to get detached from the world of things, and to stand out as a self-subsistent *tertium quid*. Instead of being an adjective of the thing we perceived, it becomes a substantive in its own right. When, for example, I try to draw that fruit bowl on a sheet of paper, I find myself seeing the orange as a flat yellow disc, and the rim of the bowl as an oval outline rather than a circular edge. It is, I am convinced, this unnoticed change in the focus of attention that accounts for the sharply conflicting explanations we meet with in all these discussions. The naïve realist keeps to the attitude of the plain man in ordinary everyday life, and his attention is therefore fixed on the object; the philosopher and the introspective psychologist fasten their attention on their sensory experiences, and the inevitable outcome is a dualistic 'sense-datum theory'.

As a scientist, the psychologist's task is not merely to describe experience, but to dissect it. In doing so he tends to analyse the perception of things present rather as he would analyse the remembrance of things past. In the end he is almost bound to adopt something like the representative theory, though I agree with Dr. Hirst that the term 'representative' is unfortunate. Such a theory can only be adequately expressed by postulating three unique *relations*—relations that are peculiar to cognitive processes as such, namely, the relations between (i) the percipient and the sensum, (ii) the sensum and the percept, and (iii) the percipient and the percept. The percipient *directly* apprehends the sensa; these sensa *imply or signify* a corresponding object; but usually there is a kind of short-circuiting, and the percipient seems *immediately* to perceive the object, and to perceive it as external. This third relation—the relation of perceiving—is, so to speak, the 'relative product' of the relations of sensing and implying.

---

[1] As Martin has observed, "ordinary common sense usage naturally embraces such facts as the relevance of the observer and his position, etc., to the apparent characteristics of physical objects" (*Sense-perception and Matter*, 1950, p. 32). Most of the older introspective psychologists adopted some such assumption: the first to reject it seems to have been James Ward. "At first sight," he says, "it looks rather as if the activity might vary, while the object remains the same." But "the theory of presentations" (his word for 'sense data') requires us to say that it is the 'presentation' that alters, not the mode of consciousness. In my view the most satisfactory way of answering the familiar epistemological puzzles (those raised by the occurrence of illusions, and the like) is provided by McDougall's version of the doctrine of apperception (cf. his *Physiological Psychology*, pp. 93f.). I tried to indicate how much of this could be formulated in terms of matrix algebra in an earlier article in *Psychometrika*, III, 1938, p. 160, 'The Unit Hierarchy'). "In psychology, as in modern physics, an 'observation' is to be specified not by a quantity but by a structure, not by the correspondence between a single sensation and a single stimulus, but by the correspondence between patterns or *matrices* . . . We may then regard the real object as specified by the canonical form of the matrices." And "if the entire matrix of stimuli, not the single stimulus, is treated as the unit, the transformations of the pattern" (e.g. the transformations that occur as I change my point of view or the object changes its distance or orientation, and the inverse transformations which my brain unconsciously and automatically carries out in the process of perception to preserve 'constancy') "are then as simple to examine as other transformations. And the postulate that introspection shall not distort what is observed can be represented by an idempotent operator".

By adopting a relational analysis instead of the traditional subject-and-predicate analysis, we avoid the erroneous inference that the sensa are 'ideas in the mind' (as the older writers maintained), or (as Russell has declared) that they are 'events in one's head'. Nor need we conclude, as Dr. Hirst has done, that, because 'perceiving' (in the sense of the physical, physiological, and psychological conditions underlying the final cognitive relation) "is a highly complex process", it therefore follows that "perceiving" (in the narrower sense of the resultant cognitive relation) must be "far removed from simple awareness". Puzzling as it may seem to the non-psychological thinker, the fact remains that in normal circumstances perception does unquestionably possess something of the sheer intuitive character of 'simple awareness'.[1]

This view, I believe, could be fully substantiated by a survey of the relevant results from experimental work on perception and attention. But even were we, with Dr. Hirst, to reject all forms of the sense-datum theory, we still could not completely escape a dualistic conclusion. Suppose, for the sake of argument, we agreed that all sensory qualities are intrinsic qualities of the physical object or its surfaces and not the attributes of certain non-physical entities called 'sense-data', we should still have to recognize the occurrence of mental images in memory, of hallucinations in delirious and psychotic states, and of all the stuff that dreams are made on; and these are unquestionably non-physical phenomena. What is more, the processes of sensing and cognizing, not to mention feeling and desiring, are from their nature non-physical. To maintain that these and other mental activities are merely the 'inner' aspects of material processes in the brain, and that the two types of process are therefore in fact 'identical' is surely no more than a verbal subterfuge. The very use of such antithetical terms as 'outer' and 'inner' is a tacit admission of their disparity.

There is one further criticism I should like to make. Throughout his account Dr. Hirst seems to suppose that, whereas the 'sense-data' assumed by his opponents must of necessity be 'private', the 'object' as such is defined by the fact that it is 'public', i.e. (as he puts it) "shared by other members of the community". Although many psychologists would agree in making publicity the main ground for believing in the existence of 'external objects', such an interpretation is to my mind quite mistaken. If, like an indigenous Robinson Crusoe, a new-born infant could survive in perfect solitude on a desert island, the child would still see the things around it—the rocks and the palms, the parrots and the goats—as objects external to himself. The real ground for distinguishing physical objects from our sensations or perceptions is that they are felt to enjoy a continuous and relatively permanent existence of their own, whereas our perceptions are of necessity intermittent, fragmentary, and subject to perpetual interruption. The primary function of the physical object is to fill in the temporal gaps; their second function is to furnish an intelligible and more or less consistent causal basis so that prediction becomes possible. They are, in the statistician's jargon, ' invariants '. Their so-called ' publicity ' is an incidental feature: indeed, the plain man attributes just the same publicity to sensory qualities as he does to physical objects.

---

[1] Cf. Burt, 'The Concept of Mind', *J. Psych. Res.*, I, ii, 1960, pp. 1–11. Although I have described the relation of perception as a 'relative product', I should still contend that in itself it is unanalysable: it is not, like the relation between uncle and nephew, *reducible* to two logically prior relations. Nor can I regard it (with some writers) as a special kind of three-term relation: e.g., "the *orange* appears *yellow* to *me*". This type of statement occurs as a natural description only in somewhat exceptional circumstances (e.g. when there is a debate about the apparent quality), and cannot be erected into a theory covering all forms of perception. It should be noted that the three statements (i) 'This object which I now see appears yellow to me', (ii) 'What I now see appears to be a yellow object' and (iii) 'The object which I now see is a yellow appearance' ('the basis of the sense-datum theory) are repeatedly treated as equivalent; actually they are very different judgements.

Most of the criticisms that I have ventured to make turn on relatively minor issues; and often, I suspect, the difference between us resolves largely into a matter of wording. The weakest point in Dr. Hirst's argument is, in my view, his attempt to remove the paradoxes of perception by invoking an Identity Hypothesis. But even here he is searching in the right direction, though what he finds is hardly the proper key to the puzzle. And he has certainly brought out a most important point that philosophical writers have been prone to overlook—namely, that the first requisite for a satisfactory solution must be a sound psychological theory based on a comprehensive survey of the relevant psychological facts. All psychological students, therefore, should be encouraged to read this illuminating survey, and consider for themselves the ingenious hypotheses that Dr. Hirst has here put forward.

CYRIL BURT

---

[1] I use, as Dr. Hirst does, the phrase ' physical object ' as the most intelligible in this context. Strictly speaking, however, that should be reserved for 'the object as conceived by physical science'; and I should prefer the term ' perceptual object'. But that would mean needless complications in my account.

BOOK REVIEWS

Individual Choice Behaviour: A Theoretical Analysis. By R. DUNCAN LUCE. New York: John Wiley & Sons; London: Chapman & Hall, 1960. Pp. xii+153. 48*s*. net.

Acts of choice, implicit if not explicit, enter into a wide variety of topics studied by the psychologist—discrimination, learning, information theory, decision theory, utility theory, and almost every form of motivation. Professor Luce's aim has been to develop a general mathematical theory applicable to all these different fields. Throughout his discussion he assumes that the situations with which the psychologist is concerned are of the following type: there is a relatively homogeneous 'universe' of alternative possibilities—it may be a specific range of stimuli, as in experiments on sensory discrimination or subjective preference, or a specific range of responses, as in studies of learning and of motivation. This 'universe' (U) is such that, out of certain finite subsets which it contains, a subject can select that particular alternative (or group of alternatives) which he considers to be the most appropriate as judged by some definable criterion (e.g. relative loudness in the case of auditory stimuli, likelihood of relative satisfaction in choosing alternative modes of behaviour). After reviewing the different types of problem arising in such situations, Professor Luce arrives at the conclusion that 'choice behaviour' may best be treated, not after the manner of previous workers as an 'algebraic phenomenon', but rather as a special type of' probabilistic phenomenon'.

His book is divided into four main chapters. The first sets forth the basic theory. Since the process involved is somewhat complex, his method assumes, for reasons that become clearer in the course of the subsequent discussion, that the final selection can be analysed as though it had been reached by stages. To take an example from a recent British experiment on artistic preferences, Miss Williams showed a set of nine pictures individually to a class of children, and required them to work by successive steps. Each child was asked first to select the three he liked best, and the three he liked least, leaving three about which he had no strong feelings one way or another. Then with each subset of three he was invited to make a similar threefold choice—i.e., to pick the one liked most and the one liked least. The ultimate result was to get the whole series of nine ranked in order of preference.

In setting out his formal proofs Professor Luce begins by enunciating what he calls the fundamental 'choice axiom'. This is expressed in the usual symbolic notation, and lays down the basic conditions of independence and transitivity. Translated into verbal terms, the first part of the axiom may be roughly interpreted as follows. Suppose that R is a small subgroup of possibilities (in a test of aesthetic preference it might be just a single picture), and that R is contained in a wider subgroup S (e.g., a batch consisting of three pictures), and that S in turn is contained in the total sample T (in our example the whole set of nine pictures): then the probability that a subject will select R out of T, namely $P_t(R)$, will be equal to the product of the probability that he first selects S out of T, i.e. $P_t(S)$, multiplied by the probability that he then selects R out of S; or in symbols.

$$P_t(R) = P_s(R)P_t(S).$$

If the various probabilities are estimated on the basis of proportionate frequencies obtained from a reasonably large number of subjects, then it is obvious that in certain somewhat trivial cases this part of the axiom would be virtually a tautology. In cases in which the selections are strictly random, the equation states the familiar theorem of 'compound probability' for two independent events. It is, however, by no means difficult to imagine certain realistic situations in which the formula might break down rather badly.

Nevertheless, intuitively at least, it would certainly seem plausible as a working hypothesis for clearly defined sets in which the number of alternatives is comparatively small. Moreover, such a restriction presents no serious drawback, since (as it turns out) Professor Luce's own deductions merely require that the axiom shall be true for sets containing at most three alternatives.

The second part of the axiom simply states that, if y is invariably chosen in preference to x, then x may be deleted from T whenever we are considering a choice from T. To take Professor Luce's example, " if I always choose liver in preference to beef, then for me the choice between the three items—beef, liver, and chicken—can be reduced to a simple choice between liver and chicken ". It is surprising to find what a number of interesting consequences, some of them quite unexpected, may be deduced from this twofold ' axiom '.

A crucial problem in the study of choice behaviour is the construction of a numerical scale which will represent the behaviour of the individuals concerned. In the past various procedures have been suggested, and provisionally adopted, by economists, by psychologists, and by psychometricians. All are open to criticism. Professor Luce therefore proceeds to demonstrate that, for pairwise discrimination, his axiom implies the existence of a *ratio scale*, i.e., a linear scale that is unique except for its units. His first step is to formulate an existence theorem for a ' local region ', i.e. for subsets of T: then, by using overlapping subsets, he shows how the ' v-scale ' (as he calls it) can be extended over the whole range of the ' universe ' of stimuli or responses with which the investigator may be concerned.

The remaining chapters—in many ways the most illuminating—deal in turn with the application of his ' basic theory ' to the commoner problems of psychophysics, of utility theory, and of learning theory. The properties of the more important models are examined; and at various points practical suggestions are put forward in regard to the best methods of estimating the underlying probabilities and the requisite parameters (e.g. by applying the method of maximum likelihood). From time to time Professor Luce discusses ' previous work '; but here his references are for the most part confined to American publications. For example, the " rather complex, normally distributed, but unobservable subworld " postulated by Thurstone as a basis for his scaling methods is described and criticized in some detail; but the improved scaling procedures based on R. A. Fisher's principles are never mentioned. There are instructive discussions of ' rank ordering ' and the method of ' paired comparison ', but no reference to the work of Kendall and his fellow-workers. However, it was not part of Professor Luce's intention to offer a comprehensive survey of alternative modes of approach, but merely to illustrate the nature and merits of his own; and this he has done with great ingenuity and skill.

In this brief outline of his general procedure I have tried to interpret it from the standpoint of British work, and to see it in the context of the problems that have interested the British investigator. I hope that in so doing I have not too seriously distorted the strict and rigorous exposition to which he himself has adhered. It is undoubtedly much to be desired that psychologists in this country will turn their attention to the aspects of the question he has raised and to the mode of solution he has advocated. He has himself, in his final appendix, summarized the more urgent of the many unsolved issues that still remain. Of these perhaps the most obvious is the need to test the validity of his basic axiom by *ad hoc* investigations, and to determine the special conditions under which it holds good. For stepwise scaling I would suggest the binary method of ' expressing as bipolar factors ' (this *Journal*, *IV*, p. 196). Still more important, though still more difficult as Dr. Luce points out, is the task of developing an adequate procedure for " describing the way organisms subdivide their environments into chunks and establishing the laws that govern the relations of one sub-division to another ". Research along these lines can hardly fail to be fruitful.

Cyril Burt

Modern Factor Analysis. By HARRY H. HARMAN. Chicago: The University of Chicago Press, 1960. Pp. xvi + 469. \$10·00.

Twenty years have passed since Professor Holzinger and Dr. Harman first published their ' Synthesis of Factorial Methods ' under the brief title of *Factor Analysis*; and during that interval the whole subject has made considerable strides. A new edition of the work had already been planned, when, to the deep regret of all his colleagues both in America and over here, Dr. Holzinger died. The new volume is something more than a mere revision. Among the novel features perhaps the most significant are the detailed accounts of methods of high-speed calculation which have resulted from the introduction of electronic computers. Analytical methods of rotation and statistical procedures for testing hypotheses are described in some detail. An entirely new ' part ', consisting of 40 pages of ' problems and exercises ' (with answers) arranged according to chapters, has now been added; and the bibliography has been enlarged to over four hundred entries, including thirty papers published in this journal.

To make room for these additions, much of the earlier volume has been omitted or abridged. And at the same time there seems to have been a change of attitude in regard to factorial techniques. With British investigators one of the commonest factorial schemes has always been that which includes first a ' general factor ' and then a set of ' group factors ', often overlapping, and not infrequently subdivided into still smaller ' group factors '. This type of pattern is now only mentioned in passing; and the account of Holzinger's own ' bifactor method ', which gave a simplified scheme of this kind, has been greatly condensed. On the other hand, the discussion of the methods and concepts associated with the ' Thurstone school ' (notably ' simple structure, which formerly was dismissed in a short footnote) has been expanded to over a hundred pages. Nevertheless, the book as a whole is admirably planned, and the exposition is in the main clear and impartial. There can be little question that the volume will at once become one of the standard textbooks on factor analysis. Indeed, in the reviewer's opinion, it is by far the best at present available.

CHARLOTTE BANKS

Information and Error. By SOLOMON DIAMOND. New York: Basic Books, Inc., 1959. Pp. xii + 307. \$5.00.

" Whoever wishes to explore the mysteries of modern psychology," says Dr. Diamond in his preface, " must carry with him an equipment of statistical techniques that were esoteric twenty years ago; and this situation calls for a new approach in the teaching of statistics." He believes that the most urgent need is a change to an ' analytic viewpoint ' as distinct from a ' descriptive viewpoint '; and it is the latter, so he maintains, that has dominated statistical teaching in psychology hitherto. The older textbooks were restricted to a study of probable or standard errors and correlational techniques; the newer textbooks tend to pass these by, and deal by preference with those procedures that are based on socalled analysis of variance. Accordingly, he has decided to include within the covers of a single textbook both the earlier Pearson-Thorndike approach (slightly modernized and supplemented with detailed discussions of Thurstonian factor analysis) and the newer methods of the Fisher-Snedecor school (supplemented with a final chapter on ' nonnormal data ') and to rewrite both from his ' analytic ' standpoint. At the same time an attempt is made to enliven the algebraic exposition with touches of American humour. The more technical section headings—' null hypothesis ', ' chi square ', ' regression formulae ', and the like—alternate with quizzical titles like ' the professor's quandary ', ' the loving sons ', and ' the tale of the scatterbrained boy '. The British student does not, as a rule, greatly relish facetious interpolations in his textbook; but he appreciates a little light relief in his lectures. And psychologists who have to discourse the somewhat arid problems of statistical techniques may find in Dr. Diamond's book much that is suggestive as regards both subject matter and treatment.

CHARLOTTE BANKS

Behaviour. By D. E. Broadbent. London: Eyre & Spottiswoode, 1961. Pp. 215. 21*s.* net.

Even the sternest critic of behaviourism as a basic psychological theory must pay his tribute to the wealth of information which behaviourism as a method has furnished about some of the most elusive problems of psychology. Indeed, owing to the constant controversies between the leading members of the school the sum total of their joint contributions has, if anything, been underrated. It is the purpose of Mr. Broadbent's book "to summarize the achievements of the behaviourist approach up to the present day". As a general review it is comprehensive, critical, and thoroughly impartial in its recognition of the merits of the rival theorists. It is full of suggestive analogies and illuminating asides, and spiced with unexpected touches of humour. The style is easy and non-technical, so that not only the student of psychology, but also the general reader can follow the arguments. The criticisms are inspired by a robust common sense, and well worth the attention even of the specialist who is already acquainted with the relevant facts.

Behaviourism was in many ways a youth movement; Watson and most of his followers were then in their early thirties. It started as a protest rather than as a fresh discovery or a novel theory. Accordingly, Mr. Broadbent begins his book with a summary of the chief defects of 'introspectionism' as seen through the eyes of Watson when he first launched his attack half a century ago. The introspective method, it was alleged, had produced nothing but conflicting results; it provided no explanation for its own findings; and it offered no satisfactory procedure for studying or interpreting the behaviour of animals. Such complaints may have been warranted in the case of American or German psychology—as exemplified by the inconclusive controversies over the James' description of emotions and the Würzburg doctrine of imageless thought (to quote Mr. Broadbent's own illustrations). But the criticisms were scarcely justified in the case of British psychology.

Long before the date of Watson's celebrated manifesto, McDougall had rejected the traditional definition of psychology as 'the study of human experience', and had described its true aim as 'the explanation of the behaviour of men and animals'. He too declared in no uncertain terms that "we cannot hope to arrive at an adequate understanding of behaviour by the methods of introspective psychology alone".[1] Years before Watson published his first paper,[2] students like myself and my contemporaries had learned from McDougall at Oxford or Myers at Cambridge the need for "objective methods of experiment", which Mr. Broadbent cites as Watson's 'central belief'. We had been warned of Lloyd Morgan's canon—"not to interpret an animal's behaviour in terms of higher quasi-human processes, when it can plausibly be accounted for by lower quasi-mechanical processes of a simpler type".[3] We had heard about the irreconcilable conclusions of different introspectionists; but we were told such conflicts were only to be expected in view of the wide individual differences which Galton himself discovered in almost every field, particularly in his studies of imagery and emotion. Stout, as well as McDougall, would have readily agreed that self-observation alone could never fully explain what self-observation might reveal. It was on this very ground that they put forward the theory of 'psychoneural dispositions', based largely on Sherrington's neurological experimental studies of elementary reflexes. From Sherrington himself we heard, long before Watson mentioned them, about the researches and speculations of Pavlov.[4] And in McDougall's theory of the learning process we encountered (couched in slightly different terms) nearly all the competing partial theories put forward by the different exponents of post-Watsonian behaviourism.

---

[1] W. M. McDougall, *Physiological Psychology*, 1905, pp. 1–5. It is curious to note that the very wording is almost identical with Mr. Broadbent's.

[2] A monograph (1907) dealing with the 'sensations' (sic!) of the rat in solving the problems of the maze.

[3] Ll. Morgan, *Animal Behaviour*, 1900, p. 270.

[4] The most important of Pavlov's early papers (*Ergebn. d. Physiol.*, III, i, 1904, pp. 177f.) is cited in Sherrington's *Integrative Action of the Nervous System*, 1906. We there learn too of the

With most of this, I imagine, Mr. Broadbent would fully agree. His purpose has been, not to write a history of psychology, but to record and evaluate the fruits of the new method of attack. What then were the novel or distinctive points in Watson's own teaching? Three main principles, we are told, stand out: the rejection of all introspection, a belief in the superiority of environment over heredity, and a simple faith in conditioning as the supreme explanatory process; and Mr. Broadbent proceeds to show how each of these principles has been progressively modified in varying degrees by the later members of the school.

To his English contemporaries Watson's earlier pronouncements seemed at first to be little more than a revival of the Spencerian type of associationism which had been so drastically criticized by Ward and Stout. It had been the claim of the British associationist school to have discovered the 'laws of association'; these, it was maintained, governed the processes of learning and therefore all forms of adaptive behaviour; and it was the cherished ambition of nineteenth century psychologists to reduce these laws to a minimum number of basic principles. Those who accepted the threefold division of mental process commonly recognized three: (i) the cognitive law of temporal contiguity (usually supplemented by a law of 'exercise' or 'frequency'), (ii) the conative principle of motivation or drive, and (iii) an affective law generally known as the pleasure-pain principle (usually expressed in terms of 'satisfaction' or 'dissatisfaction' which were supposed to reinforce or eliminate the 'associations' they accompanied). Watson discarded the last two on the ground that they attributed a causal agency to consciousness, and retained merely what Broadbent calls "the conception of stimulus-response association by contiguity". The term 'conditioning' was used to designate the hypothetical process of 'switching', whereby a new reflex stimulus or a new reflex response was (in Watson's phrase) 'substituted' for the first.

However much we may criticize the details of Watson's teaching, we must give him full credit for the widespread enthusiasm his exhortations inspired. But during the next ten or fifteen years a vast accumulation of experiments, carried out by an army of independent workers, gradually made it clear that his extremely simple hypothesis left a wide variety of results wholly unexplained. And it became obvious, as Broadbent demonstrates by convincing instances, that "Watson's neglect of motives and rewards had gone too far". At length, in the early thirties, another theory of learning was put forward, with the 'law of effect' as its central principle. The 'effect', however, of reward or of punishment was now treated as purely mechanical.

This change in standpoint is generally associated with the name of Clark Hull, a comparatively late recruit to the behaviourist fraternity. His aim, as he tells us, was "to reach a quantitative determination of the laws of behaviour and to arrange them in a deductive system". The underlying theory, formulated with all the logical rigour of an 'axiomatic procedure' (rather like Spinoza applying Euclid's methodology to problems of human conduct) is in its basic nature very similar to Lloyd Morgan's account of habit-formation through trial and error. Broadbent here draws a most instructive parallel. The doctrine of natural selection maintains that new species are evolved as a result of chance variations in individual structure, among which only the best adapted survive; the law of effect maintains that new habits are acquired as a result of chance variations in transitory reactions, out of which none but the most appropriate are preserved. And he drops a suggestive hint for the statistical psychologist. "Curiously enough," he says, "nobody seems as yet to

importance of conative *driving force* and of affective *reinforcements*, and how reflexes may be modified by the coordination of one reflex arc with another. However, we talked of 'associated reflexes', not of 'conditioned reflexes', so there was as yet no popular slogan to publicize the new and fruitful procedures devised by the Russian physiologists.

have applied to animal behaviour the mathematical arguments used in dealing with the evolution of species."

However, still further experiments quickly revealed three types of case in which the results observed seemed quite inconsistent with the new doctrine, viz., " first, failure to learn even when a reward was given; secondly, successful learning when no reward was given; and thirdly, the learning of a mode of action different from the action that had been rewarded ". There is one characteristic type of behaviour observable in the rat—the animal most frequently studied by psychologists in their work on learning: to most of us it is familiar from the habits of our domestic pets, namely, their use of exploration as a means of spontaneous learning. Put a hungry rat into a strange maze. Its first reaction is, not to run straight for the food that it can smell or see, but to inspect and investigate its new quarters. It thus acquires, as a few simple tests will show, a systematic knowledge of its surroundings. There is here no question of punishment or reward, unless with McDougall we regard a satisfied curiosity as its own reward. The non-psychological reader will think there must have been some kind of blind-spot in the eyes of the earlier observers, if (as Broadbent assures us) so conspicuous a mode of behaviour was ignored for thirty or forty years. The advantages of this instinctive propensity can readily be understood in terms of a new concept put forward by Tolman. His own name for it was ' sign-Gestalt-expectation ', which Broadbent neatly translates as ' the map-in-the-head theory '. Tolman, so Gardner Murphy remarks, " agreed with McDougall " (and it might be added, with biological investigators like Tinbergen and Lorenz) " in emphasizing the need for a *purposive* approach: the organism perceives objects as means towards goals ". The most distinctive feature in Tolman's new approach consists in the fact that he asks us to turn our attention, not so much to the motor or response-aspect of the total process, but rather to the cognitive or perceptual aspect.[1] In its generalized form it is plainly capable of accounting for many phenomena which defied interpretations in terms of the older and simpler doctrines.

One of the chief reasons given by Watson for rejecting the ' subjective ' and ' anthropomorphic ' terminology of the introspectionist school was (in Broadbent's phrase) that " it made comparison with animals difficult ". At the same time the behaviourists' virtual rejection of hereditary factors has led to " a neglect of innate behaviour " and a concentration on " the mechanisms of learning " as studied in the laboratory. If, however, one extends one's observations from behaviour in the laboratory to more natural conditions, and from the humble rat to the higher carnivores and primates, or even to the domesticated cat and dog, then, I am tempted to say, the older mode of approach really makes comparison easier. There is a valid as well as an illegitimate type of anthropomorphism; and the psychological concepts we use should be applicable to our fellow-mammals as well as to our fellow-men. This was Darwin's attitude; and he for one felt no scruple over speaking about " emotions in man *and animals* ". " Higher vertebrates," says Julian Huxley, " have behaviours similar to ours in many essential ways: why deny them psychological experiences similar to ours? To give the fullest possible interpretation to their behaviour we must have recourse to a language that employs a subjective as well as an objective terminology."[2] Mr. Broadbent, I imagine, would largely agree. He too maintains that " quite complex actions may be necessary to the animal just as primary or physical ' needs ' are—a fact which a psychologist studying rats in mazes might never have discovered ".

---

[1] Tolman himself relates that he " borrowed the notion that purpose and cognition are essential ingredients from Perry "; and his own efforts were largely concerned with " translating familiar pre-behaviour concepts, such as ' conscious ', ' emotion ', ' perception ', etc., into new, non-physiological behavioural terms . . . free from introspective definitions " (*History of Psychology in Autobiography*, IV, 1952, pp. 332f.).

[2] Introduction to Joy Adamson's *Living Free* (1961), pp. x-xi.

Broadbent goes on to emphasize a further weakness in the work of these earlier theorists: they seem to be not only out of touch with the relevant work carried out by contemporary biologists (notably those who were studying instinctive behaviour); they also showed " a curious lack of any connection with physiology". And equally, he says, the textbooks of physiology evince a curious ignorance of recent psychological work: " they still give Pavlov's interpretation as though it had remained unquestioned ". A strenuous attempt to overcome this defect was made by Lashley: every student knows how his experiments on rats and other animals threw the gravest doubts on the old concept of cerebral localisation. " The first impact of this finding," says Broadbent, " was to encourage the view that regarded only patterns as crucial." In consonance with the earlier speculations of Köhler and his fellow Gestaltists, " the incoming messages from the senses were now thought of as setting up fields of electrical potential on the cortex ".

But meanwhile evidence was rapidly piling up to show that what Adrian had called " the all-or-none principle in nerve-conduction " holds good, not only of afferent and efferent neurones (where it can readily be demonstrated), but also of cortical neurones. This conclusion seemed hardly compatible with the electrical field-theory, at any rate in its simple form; and later experiments by Lashley and others " cancelled out the boost which his earlier experiments had given the theory ". Hebb therefore suggested an alternative picture for the way behaviour is ' organized '. To me, I confess, his descriptions of ' cell assemblies ', ' neural after-discharge ', and ' phase sequences ', instructive as they are, appear little more than a schematized version of the old theory of the ' integrative action of the nervous system' particularly as it was developed by McDougall in his account of how ' psychoneural dispositions ' are formed at the higher levels—though, of course, Hebb's formulation is expressed in a novel and more up-to-date terminology. But, as Skinner and other critics were quick to point out, it still remains almost entirely hypothetical.

At the present time, we are told, enthusiasm for these various extensions of the original behaviouristic doctrine is on the wane. Today the methods most frequently used, and the interpretative principles most widely adopted, are those of Skinner, the youngest of the behaviourists we have so far mentioned. In many research laboratories the ' Skinner box ' has almost entirely ousted the manifold versions of the rat-maze. Skinner, like Thorndike of old, and like Mowrer, Maier and Schneirla, and several other contemporary writers, overcomes one of the long disputes which divided the earlier behaviourists by recognizing that both the rival theories contain an element of truth. There are, he maintains, *two* basically different types of learning process—by mere ' contiguity ' and by ' reinforcement ' (or ' effect ') respectively, though he gives each of the traditional principles a slightly novel twist. In Skinner's terminology the former is called ' respondent ' or ' classical conditioning ', and the latter ' instrumental ' or ' operant conditioning '. One of his special devices is to programme the method of learning by subdividing it into elementary steps, each with its own reward, so that the creature learns the final form of behaviour without any needless error—a principle more recently embodied in his ' teaching machines ' for human pupils.

Above all, Skinner has endeavoured to adhere strictly to the principles of ' operationism ' (in Bridgman's sense) and to eliminate the various ' un-point-at-able ' structures and processes that had crept back into the later explanatory theories. Instead, he seeks to discover ' invariable deterministic ' laws, free from all contamination with ' chance factors '. However, as Mr. Broadbent justly points out, a methodology of this strict and narrow type is almost bound to leave out much of what we really know, or can plausibly infer, about behaviour and its underlying mechanisms. He criticizes in some detail Skinner's " hostility to statistics "; after all, " the *variability* of behaviour is one of its most crucial features ".

Most of the foregoing developments, which Mr. Broadbent has reviewed at length, and I have epitomized so far as possible in his own words, will already be familiar to the reader. But it is precisely his way of wording or rewording these familiar details which is to my mind so instructive. Often by a novel turn of phrase he succeeds in bringing out a fresh aspect in what was otherwise well known; and at every point his explicit criticisms are based on an unbiased judgement and sound practical sense. His applications of the results to human behaviour are always enlightening.

His last chapter looks back over the field, and sets out his final conclusions in regard to the behaviourists' achievements and their chief weaknesses and flagrant omissions. " Facts in profusion have been discovered, some surprising, but most of them in accordance with traditional wisdom "; but the outstanding merit of the movement has been to provide " objective means of testing intuitive knowledge ". Its most obvious defect is that " almost all the research has been concentrated predominantly on the rat." Most of it has been laboratory research. Unlike the zoologist, the behaviourist rarely starts by " observing an animal in its natural surroundings ". " Another besetting weakness of behaviour theory has been . . . its tendency to premature formalization". The majority of theorists, notably Clark Hull, have put too much faith into the ' hypothetico-deductive procedure ' (or, as Mr. Broadbent dubs it, ' the guess-and-check method ') which, as he repeatedly reminds us, commonly resulted in a neglect of other possibilities. " Alternative theories should be set up, instead of putting forward just one guess alone." A promising way for overcoming these limitations would be " to use more sophisticated mathematics: with mathematical reasoning it would be impossible to make false inferences or to rely on concealed assumptions to such an extent as theories of behaviour have done in the past ". The statistical investigations of Estes, and of Bush and Mosteller, are cited as illustrating what may be attempted along these lines. But the mathematical analysis should be " based on experimental work of comparable rigour and scope ".

" Perhaps the worst weakness of behaviour theory is its method of dealing with events inside the brain, revealing themselves only indirectly in action." The recent theoretical work on electronic machanisms which carry out guidance and control, or undertake the analysis and transmission of information, might, he believes, provide a useful descriptive terminology—a view developed with much skill in his earlier volume. I am tempted to make one further suggestion. Time after time the development of behaviour theory, as we have seen, has proceeded by reintroducing, under new and non-committal names, concepts in regular use among psychologists of the introspective, pre-behaviouristic school. Many of those who have studied the problems of learning among human beings—particularly among children—would be inclined to urge that the ban on introspective checks and on theories frankly expressed in introspective language should now be lifted. It is quite as easy to check the results of self-observation as to check the results of observing rats. Mr. Broadbent himself contends that " it is the quality of his experience which makes Man so valuable "; and (if I understand him rightly) he would not deny that much of this can be " communicated in the language of science ". Indeed, he turns the tables on those critics of behaviouristic methods who hold that such doctrines tend to " undermine morality and responsibility " by declaring that " if we were to reject the scientific approach because it is not private, we should also be rejecting all attempts to share the experience of others, whether artistic, religious, or of some other kind ": the very attempt to communicate such experience " must rely ultimately on the description of situations and actions ". After all, he asks, " is it not egocentric to have no curiosity about the richness and variety of other people ? "

CYRIL BURT

The Language of Psychology. By GEORGE MANDLER and WILLIAM KESSEN. New York: John Wiley & Sons. London: Chapman & Hall, 1959. Pp. xviii + 301. 54*s*. net.

" This book," as the authors explain, " is intended to be a compound of the analytic and behavioural view of science," and with this as their twofold guiding principle they have sought to " provide the student with an introduction to the language of science in psychology and the social sciences." The analytic standpoint is that of American logicians, like Quine and Hempel, rather than that of the earlier British champions of linguistic analysis; and the behaviourist standpoint is that of Hull and Skinner rather than that of Watson and his immediate followers.

The book is divided into two main parts. The first deals with ' the vocabulary of psychology '. The idea of a " psychological vocabulary rooted partly in immediate phenomenal experience " is rejected on the familiar ground that experience is private and not public. The study of perception is not absolutely banned, since after all science is based on perception; but we are told that it must " be dealt with in terms of verbal reports ". After a still more detailed consideration the authors also decide that neither a " commonsense thing-language " nor a " purely physicalistic language " would be sufficient by themselves. In the end they conclude that a " basic criterion " is needed to decide whether or not a word or phrase is admissible as a " scientific term "; and they maintain that the only satisfactory test is " response invariance "—a description which covers precision, reliability, and consistency as regards both ' intra-subjective ' and ' inter-subjective ' usage. The term ' response ', be it noted, includes internal or ' organismic ' responses as well as overt or external responses.

The second part is entitled ' theory in psychology ', and deals with the general nature of scientific theories, and with the characteristics, sources, and uses of explanatory systems. This leads to a full discussion of both induction and deduction as methods of scientific proof. ' Intervening variables ' and ' disposition terms ' are permitted; but they must be rigorously defined by means of ' reduction sentences '. However, anything like a thoroughgoing reductionism in psychology, such as was advocated by Carnap, is deemed impracticable. The volume ends with an interesting chapter on ' special problems in the language of psychology ', including a short but excellent account of scientific ' models and analogies '.

It will be seen that the criterion which the authors propose is capable of a fairly generous interpretation. Nevertheless, to insist on a vocabulary that is primarily behaviouristic—to exclude all " entity-characteristics except those for which there are theoretical or proto-theoretical physical justifications "—seems to be needlessly restrictive. It leaves out too much. As several of his critics have argued, " to invoke (as Watson did) ' internal ', ' implicit ', or ' organismic ' responses which by hypothesis are unobservable is far more precarious than frankly admitting introspective descriptions, which, in principle at any rate, can be readily checked by the observations of others; and to treat these introspective descriptions as no mere ' verbal responses ' is to divert attention to the wrong problems—to what the subject says instead of what he experiences ". After all, provided our subjects are adequately trained, their most illuminating contributions are the processes they observe, not the manner in which they communicate their observations.

However, in spite of these minor defects, the book as a whole undoubtedly succeeds in bringing together in a most helpful way the chief points made by the modern school of linguistic analysis, and in showing how they bear on the task of the psychologist. It is instructive to discover that, as a result of their own critical and impartial survey, and almost (it would seem) in spite of themselves, the requirements which the authors finally lay down are decidedly more elastic than those of the behaviourists and analysts whose somewhat dogmatic pronouncements provided the main stimulus for their work. At times their own language is liable to surprise the reader by the way their metaphorical vocabulary gets mixed: (e.g., " this curtain-raiser to the new psychology, against the background of Helmholtz's and Fechner's work on vision and psychophysics, has remained a major millstone of modern psychology "). Nevertheless, in spite of the abstract nature of the subject, the discussion as a whole is clear, readable, and fully illustrated with apt examples.

M. Howard

An Introduction to the Principles of Psychology. By B. R. BUGELSKI. New York: Rinehart & Co., 1960. Pp. xviii + 571. $6.00.

" This text," the preface tells us, " has been prepared for the qualified and serious student: it is doubtful that a smattering of ignorance should be presented to the non-major." From this the reader will gather that Dr. Bugelski's rather weighty volume is typical of the many introductory textbooks intended for the American classroom. Psychology is defined as " the scientific study of behaviour, internal and external, of organisms ". Because of their ' subjective connotations ', the use of terms like ' experience ' and ' mind ' is deprecated. The general standpoint is that of J. B. Watson as developed by C. L. Hull and more especially by B. F. Skinner. In the opening chapter Skinner is quoted as claiming that " by the proper use of psychology all students can earn A grades if properly taught: delinquency, disorderliness, and undesired modes of behaviour can be eliminated "; and the student is urged to think of psychology as a profession like medicine, and not merely as a science.

In the body of the work, however, the behaviourist views are put forward in a more tentative and flexible fashion than is customary with writers belonging to the same school of thought. When it comes to summing up arguments on such controversial issues as ' nature or nurture ', ' the evidence for extrasensory perception ', and the like, the conclusions are frequently phrased in the form of rhetorical questions rather than dogmatic affirmations, or expressed in terms of ' might be ' or ' may be '. Indeed, without the help of the ' instructor ', the attentive reader will often find himself in the dark as to what exactly is the final upshot. There are several curious errors in the spelling of names; the familiar doctrine of the four temperamental types, for instance, is ascribed more than once to ' Hypocrates '.

The student is introduced at an early stage to the elements of statistical psychology, and a chapter is devoted to ' psychometrics '. But here and elsewhere the views of older investigators—Galton, Binet, or Spearman—are not always quite correctly stated. Spearman, for example, is said to have maintained in his 1904 article that " intelligence must be thought of as consisting of a general component, G, and some unknown number of special aptitudes, S ": (Spearman, of course, denied the existence of special *aptitudes*, and identified ' intelligence ' solely with G). Much is made of the fact that intelligence as tested varies with the individual's scholastic and social environment; but, when we learn that the tests that the author has in mind " rely heavily on vocabulary items " and " call for items of information ", such a result is hardly surprising. However, the final conclusion is that " intelligence *may be* an intervening variable ", and " it *might be* that inheritance sets some kinds of limits " (reviewer's italics).

One attractive feature is the accompanying booklet which contains a series of ' new type questions ' on the topics of the several chapters, so that the student can test his knowledge and check his answers. A ' work book ' is promised which will " incorporate as much as possible of the ' teaching machine ' design ".

J. M. DAVIS

Behaviour Genetics. By JOHN L. FULLER and W. ROBERT THOMPSON. New York: John Wiley & Sons. Pp. x + 396. 72*s*.

It is remarkable that no psychologist has hitherto attempted a comprehensive synopsis of recent work on the facts and theories of mental inheritance. Biologists, from the time of Darwin and Galton onwards, have consistently maintained that heredity has much the same influence on the physical bases of mind as it has on other physical characteristics. Psychologists, on the other hand, have drifted away from the views their predecessors held, and today for the most part seem to have decided that genetic conditions can have little or nothing to do with determining the mental differences between one individual and another. The reason for this 'anti-heredity bias', so Dr. Fuller and Dr. Thompson believe, "springs largely from the social and political implications of the subject". But it has surely been also encouraged by the increasing influence of behaviouristic doctrines during the last two or three decades.

Even the authors of this admirable treatise have not escaped scot free from the bias imparted by the prevailing trends. Professor Thompson is a psychologist, Dr. Fuller a geneticist. Both have carried out original researches on the genetics of animal behaviour; and this still seems to be their dominating field of interest. Indeed, when embarking on their task, they "considered omitting human genetics completely": and, as they explain in their preface, they have deliberately avoided such old fashioned phrases as 'genetic *psychology*' or '*psychological* genetics'. Their firm resolve to describe their findings solely in terms of 'behaviour' has resulted in several surprising omissions and a number of curious circumlocutions. For instance, to treat partial colour-blindness and similar conditions as primarily forms of 'behaviour' is not merely awkward, but at times actually misleading. This mode of approach may be natural enough in studying animals; but it certainly detracts from the value which the authors' discussions might otherwise have for students of human beings. Nevertheless, in spite of its limitations, the work as a whole will prove a most welcome addition to every psychological library. J. CONWAY

Forces and Fields. By MARY B. HESSE. London: Nelson & Sons, 1961. Pp. x+318. 35*s*.

As its subtitle indicates this volume is essentially a historical study of the notion of action at distance. The hypothetical models adopted by scientists to account for the various types of phenomena falling under this head are of two main kinds: those that accept the notion of a force acting on an object across an empty interval of space, and those that require continuous transmission by a medium or field. To a large extent the investigator's preference for one kind rather than the other is determined, not so much by his empirical observations as by his tacit preconceptions as to the rôle of theories and hypotheses in scientific explanation. Accordingly, Dr. Hesse begins her book with a preliminary disquisition on 'the logical status of theories', and a detailed examination of 'operationism', the 'hypothetico-deductive method', and the use of scientific 'models'. Then in the body of the work she proceeds to trace the history of her subject from the early speculations of the ancient Greeks to those of the modern quantum physicist; and the book concludes with a chapter on the 'metaphysical framework of physics'.

The whole discussion, which is admirably documented, should be of the greatest interest to those psychologists who are concerned more especially with the methodological background of their general investigations. But there is a more particular reason why the paradox of action at a distance should appeal to the student of the mind. Francis Bacon in his classification of phenomena falling under this general head includes not only those in which the thing acted upon is a material substance or body, as in gravitation or magnetism, but also those in which it is a mind, as in the awareness of sounds, of warmth, and of visible objects, and in the communication by telepathy between friends or enemies or the influence of masterful personalities. Normal cognition, it will be noted, comes under this second category as well as paranormal. Sense-perception, as Burt has repeatedly argued, is essentially a mode of action at a distance. Dr. Hesse does not deal in any detail with normal cognitive processes; but she devotes several pages to the consideration of extrasensory perception.

Professor Broad, writing many years ago, put forward nine 'basic limiting principles' to which he believed all Western scientists would subscribe, and which appear *prima facie* to be contradicted by alleged ' psi-phenomena '. Of these methodological requirements, several, as Dr. Hesse points out, would no longer be endorsed by contemporary physicists; others raise difficult philosophical issues; and as to the remainder it would, she says, appear quite possible to devise a hypothetical mechanism which would allow us to accept such phenomena without violating any ' basic principles '. Nevertheless, she is inclined to agree that " some new explanatory concepts are needed "; and, although hitherto physical theories have proved somewhat unhelpful in suggesting working models to account for paranormal phenomena, it must still be to the conceptual systems of the physical sciences (she believes) that the psychologist will probably have to turn in the first instance for suggestive analogies. The concept of ' field ' is itself one of the most obvious. At bottom however the whole problem is not physical, but metaphysical, and, as she puts it in her closing words, " a society that is uninterested in metaphysics will have no physical science ".

T. R. GREGORY

The Achieving Society. By DAVID C. MCCLELLAND. New York: Von Nostrand, 1961. Pp. xvi + 512. $7.95.

Dr. McClelland, who is professor of psychology in the Department of Social Relations at Harvard University, is already well known in this country as the editor or co-author of works on motivation and kindred topics—notably *Talent and Society* and *Personality: The Achievement*. And the present volume is in many ways a sequel and a supplement to these two earlier publications. His problem is an ancient one. Why have certain societies at certain epochs produced so many great historical figures—statesmen, scientists, men of letters, men of business, while other societies have barely managed to survive? And why have the successful civilizations of the past so often ended in a decline and fall? The hypothesis he sets out to test is that the achievements of an ' achieving society ' are the consequences of its ' need for achievement '.

Three methods of investigation have been attempted based on (1) group measures of the ' need to achieve ' and other psychological variables, (2) individual measures of motives, values, interests, and the like, and (3) a more intensive study of the motives and behaviour of ' business entrepreneurs '. Free use is made of the analysis of stories, myths, literature generally, as well as of answers to systematic questionnaires. The results are compared with whatever data were available in regard to the economic development of the particular country in question during the particular period selected. The societies studied include those of ancient Greece, Spain in the Middle Ages, pre-Inca Peru, the United States, England under the Tudors, and during the industrial revolution and the Wesleyan revival, and (more briefly) Germany, Italy, Poland, Turkey, India, and Japan.

Dr. McClelland's final conclusions are highly instructive. The main ' sources ' of the ' need to achieve ' are in his view environmental, more particularly a temperate climate, certain cultural traditions, certain child-rearing practices, a middle class status, and a religious background of a quasi-protestant quasi-mystical type. Little attention is paid to genetic factors. Galton's eugenic theory is never mentioned. Dr. McClelland's chief example of ' biological explanation ' is a theory which he ascribes to Fanfani (an Italian politician of the thirties) who, as Dr. McClelland puts it, " decided that the shapes of their heads must be the determining factor ".

The book as a whole is a monument of industry. It owes much to the teams of advisers, assistants, and research associates who took part in many of the investigations or passed on their specialized knowledge of certain fields. Even readers who may be inclined to question the value of much of the quantitative work will nevertheless find the factual details and the deductions based on them as interesting as they are enlightening.

M. R. JOHNSON

The Logic of Personal Knowledge: Essays presented to Michael Polyani on his Seventieth Birthday: By Paul Ignotus and other contributors. London: Routledge and Kegan Paul, 1961. Pp. xii + 248. 40*s*. net.

Michael Polyani is one of the most versatile of scientists. He was trained in medicine in Budapest, and later became a member of the Kaiser Wilhelm Institute of Physical Chemistry. Since coming to this country he has held two chairs at the University of Manchester, the first in Physical Chemistry and the second in Social Studies. His earliest book was on *Atomic Reactions*. To psychologists he is probably best known by his more recent works on *The Study of Man* and *Personal Knowledge*. The essays which compose this extremely readable Festschrift range over the whole scope of his multifarious interests. They start with technical papers on physics and chemistry, and proceed in turn to problems in the history of science, and the study of society, and conclude with certain semi-philosophical issues raised by the latest developments in chemical biology.

The contributions in these last three sections will have special appeal for psychological readers. The first is a short but fascinating paper by Arthur Koestler on ' Kepler and the Psychology of Discovery ': it traces the strange and circuitous path by which Kepler, as his inquiries progressed, passed stage by stage from the traditional type of cosmology that had survived for so long—a farrago of Platonic speculations, theological symbolisms, and animistic fantasies—to what is essentially the attitude of the modern scientist; and it describes, with apt quotations from Kepler's own voluminous writings, how he welded into a single dynamic whole the disconnected fields of sky geometry and terrestrial physics, and thus laid a firm foundation for the Newtonian universe. Mr. Koestler's paper is followed by a brief historical study by Miss C. V. Wedgwood on ' The Scientists and the English Civil War ', in her usual delightful and informative style, tracing a similar change of outlook that overtook this country during the twenty-seven years that intervened between Charles I's charter for the *Musaeum Minervae* and Charles II's charter for the Royal Society.

The next section contains a series of papers on problems from the borderland between social psychology and the related disciplines of economics, politics, and law. Here one of the most interesting contributions is that by Raymond Aron, comparing and contrasting the leading ideas of Polyani and Max Weber. The book concludes with three biological papers—on the logic of biology by Dr. Marjorie Grene, on the origin of life on earth and elsewhere by Professor Calvin, and an ingenious statistical inquiry by Professor Wigner, who attempts to estimate ' the probability of the existence of a self-producing unit '. He concludes his paper (and the book) with " a firm conviction that the existence of special ' biotonic laws ' stems from the overwhelming phenomenon of consciousness ".

Cyril Burt

INDEX TO VOLUME XIV

ARTICLES

Author	*Title*	*Page*
Bolton, T. H.	The Structure of the Mind, II	144
Burroughs, G. E. R. and Miller, H. W. L.	The Rotation of Principal Components	35
Burt, C.	Intelligence and Social Mobility	3
Burt, C.	Factor Analysis and its Neurological Basis: A Reply	53
Burt, C.	The Gifted Child	123
Burt, C.	The Structure of the Mind: A Reply	145
Dingman, H. F., Eyman, R. K. and Miller, C. R.	Factor Analysis, Latent Structure Analysis and Mental Typology	29
Eyman, R. K., Miller, C. R. and Dingman, H. F.	Factor Analysis, Latent Structure Analysis and Mental Typology	29
Gibson, W. A.	An Asymmetric Approach to Multiple Factor Analysis	97
Gregson, R. A. M.	The General Aesthetic Factor, I	72
Harrison, A. L.	High Ability and High Achievement	171
Hudson, K. S.	The Structure of the Mind, II	143
Jarrett, R. F.	A Note on Rotation and Psychological Space	109
Kaiser, H. F.	A Note on Guttman's Lower Bound for the Number of Common Factors	1
Levy, L. R.	Factor Analysis and its Neurological Basis	51
Miller, C. R., Eyman, R. K. and Dingman, H. F.	Factor Analysis, Latent Structure Analysis and Mental Typology	29
Miller, H. W. L. and Burroughs, G. E. R.	The Rotation of Principal Components	35
Sandon, F.	The Means of Sections from a Normal Distribution	117
Smythies, J. R.	The Structure of the Mind, I	141
Somerville, J. F.	High Ability and High Achievement	171
Stone, M.	Subjective Discrimination as a Statistical Method	25
Williams, E. D. and Burt, C.	The General Aesthetic Factor, II	75

BOOK REVIEWS

Author	*Title*	*Page*
Broadbent, D. E.	Behaviour	184
Bugelski, B. R.	An Introduction to the Principles of Psychology	190
Cohen, J.	Chance, Skill, and Luck	83
Diamond, S.	Information and Error	183
Downham, J.	Attitude Scaling	85
Fuller, J. L. and Thomson, W. R.	Behaviour Genetics	190
Guest, P. G.	Numerical Methods of Curve Fitting	84

BOOK REVIEWS—*contd.*

Author	*Title*	*Page*
Harman, H. H.	Modern Factor Analysis	183
Heisenberg, W.	Physics and Philosophy: The Revolution in Modern Science	94
Hesse, M. B.	Forces and Fields	191
Hirst, R. J.	The Problems of Perception	173
Husén, T.	Psychogical Twin Research	79
Husén, T.	Abilities of Twins	96
Ignotus, P.	The Logic of Personal Knowledge	193
Lange, O.	Introduction to Econometrics	86
Lipset, S. M. and Bendix, R.	Social Mobility in Industrial Society	87
Luce, R. D.	Individual Choice Behaviour	181
Maccoby E. E. Newcomb, T. M. and Hartley, E. L.	Readings in Social Psychology	90
Mandler, G. and Kessen, W.	The Language of Psychology	188
McClelland, D. C.	The Achieving Society	192
Medawar, P. B.	The Future of Man	91
Waters, R. H. Caldwell, W. E. and Rethlingshafer, D. A.	Principles of Comparative Psychology	93

INDEX TO VOLUMES I TO XIV

An index to the first fourteen volumes of the *British Journal of Statistical Psychology* classified according to authors and subjects, is now available. As proprietors of the journal the British Psychological Society wishes to take this opportunity of expressing their gratitude to Mrs. Beatrice Warde to whose generosity and expert assistance the compilation and printing of the index is due.

OCCUPATIONAL PSYCHOLOGY

Editor: ALEC RODGER

JANUARY and APRIL, 1961 — VOLUME 35, Nos. 1 and 2

Shift Changes and Hours of Work. By J. WALKER

Nurses' Ratings of Patient Welfare as Criterion Measures in the Health Sciences. By J. RICHARD SIMON

Psychopathology and Occupation: Part I. Economic Insecurity. By S. A. LEAVY and L. Z. FREEDMAN

The Socio-technical Context of Industrial Inspection. By L. F. THOMAS and A. E. M. SEABORNE

Industrial Fatigue. By E. G. CHAMBERS

Attitudes to the Employability of Chronic Schizophrenic Patients. By J. K. WING

The Psychologist's Role in the Development of Man-Machine Systems. By C. CAMERON and K. G. CORKINDALE

A Validation of Qualification Requirements for Work in a USAF Specialty. By C. J. JUDY

A Note on Skill. By B. N. KNAPP

Book Reviews.
Other Books Received.

Annual Subscription **40** *shillings*

National Institute of Industrial Psychology, 14 Welbeck Street, London, W.1

15 MAR 1962

BRITISH JOURNAL OF STATISTICAL PSYCHOLOGY

Vol. XIV Part 2 November 1961

CONTENTS

		Page
Gibson, W. A.	An Asymmetric Approach to Multiple Factor Analysis	97
Jarrett, R. F.	A Note on Rotation and Psychological Space .	109
Sandon, F.	The Means of Sections from a Normal Distribution	117
Burt, C.	The Gifted Child	123
Notes and Correspondence		
	The Structure of the Mind	141
	High Ability and High Achievement . . .	171
Critical Notice	The Problems of Perception	173
Book Reviews	. . . .	181
Index to Volume XIV	. . . .	194